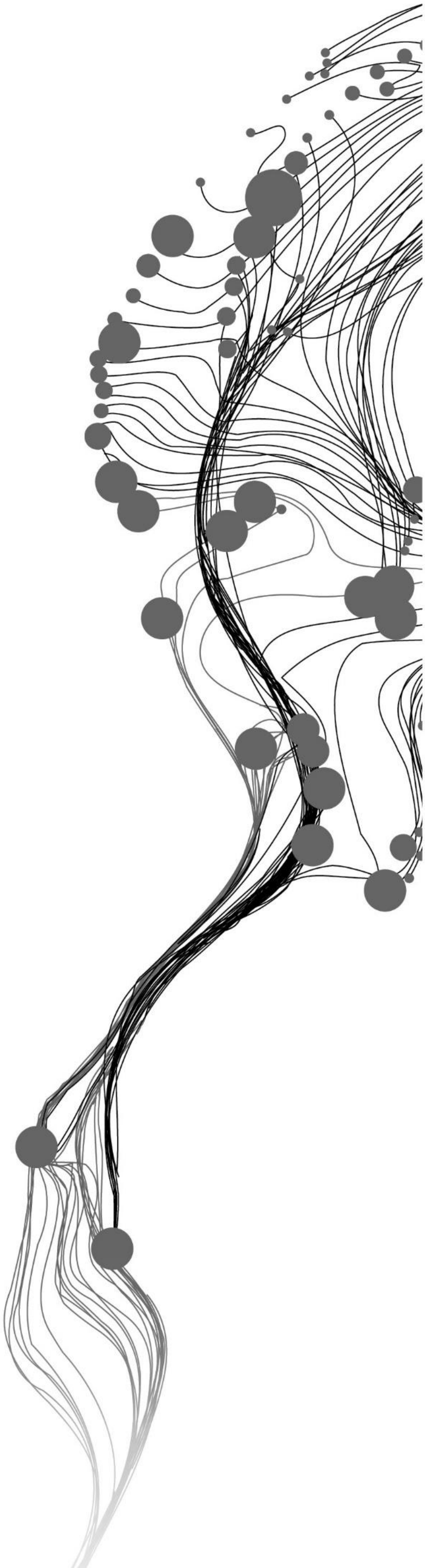
LODGING DETECTION IN WHEAT: A MULTI-SENSOR APPROACH USING SENTINEL-1 AND SENTINEL-2

SHAWL MENGISTU ASHENAFI
June 2021

SUPERVISORS:
Prof.dr. A.D. Nelson
Dr. R. Darvishzadeh



LODGING DETECTION IN WHEAT: A MULTI-SENSOR APPROACH USING SENTINEL-1 AND SENTINEL-2

SHAWL MENGISTU ASHENAFI
Enschede, The Netherlands, June 2021

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Natural Resources Management

SUPERVISORS:
Prof.dr. A.D. Nelson
Dr. R. Darvishzadeh

THESIS ASSESSMENT BOARD:
Dr.ir. A. Vrieling (Chair)
Dr. M. Boschetti (External Examiner, National Research Council of Italy, Institute for Electromagnetic Sensing of the Environment.)

ABSTRACT

Lodging is the permanent dislocation of the crop stem and root from its original position, and it affects grain yield and quality. Lodging monitoring is essential for reducing yield loss, avoiding knock-on effects, and maintaining grain quality. Remote sensing (RS) based lodging monitoring would help to acquire precise and continuous spatiotemporal data; it could help farmers increase their productivity and ensure sustainable agricultural production. In this study, the capability of data from Sentinel-1, Sentinel-2 and their combination were explored for wheat lodging detection and classification using a Random Forest classifier. Backscatter ($\sigma°$) and spectral reflectance of individual bands were extracted from the Sentinel-1 and Sentinel-2 time series data, respectively. Then the temporal backsccatter and spectral behaviour of different wheat lodging score (LS) groups (healthy (He), moderately lodged (ML), and severely lodged (SL)) were explored at different wheat growth stages via statistical analysis. A Random Forest classifier was used to classify the Sentinel-1, Sentinel-2, and their combination data into different LS classes. The statistical analysis of the Sentinel-1 backscatter and Sentinel-2 spectral data identified relationships with the field based-LS. The combination of Sentinel-1 and Sentinel-2 data based Random Forest model provided a better (83%) classification accuracy than Sentinel-1 (79%) and Sentinel-2 (80%) alone. The result implies that both Sentinel-1 and Sentinel-2 datasets provide complementary information for the model. Cross-polarised backscatter ($\sigma VH°$) was the most important variable in the Random Forest classification of Sentinel-1 data. The red-edge-1 (RE-1) spectral band was the most important variable in the Random Forest classification using Sentinel-2 data. Furthermore, $\sigma VH°$, short wave infrared-1 (SWIR-1), and RE-1 were the top three most important variables in the Random Forest classification when Sentinel-1 and Sentinel-2 were combined. Although the backscatter and spectral features of Sentinel-1 and Sentinel-2 could distinguish wheat lodging effectively, the combination of the two datasets would help to improve the classification accuracy. Therefore, applying the combinations of high spatiotemporal resolution SAR and optical remote sensing data in lodging monitoring can help reduce crop production loss and achieve higher crop yield quantity and quality.

ACKNOWLEDGEMENTS

First and foremost, praises and thanks to the holy God, the Almighty, for his support, care, and endless love during all my works and stay in the Netherlands.

I would like to forward my heartfelt appreciation and gratitude to my first and second supervisors, Professor Andy Nelson and Dr. Roshanak Darvishzadeh, who provided me valuable advice, encouragement, and critical comment during the entire research work. They have provided me unreserved support, follow-up with encouragement, provision of insightful ideas, and constructive inputs for my MSc research work. I am extremely grateful and thankful for your support and expert guidance in my entire MSc research work.

I would like to extend my sincere gratitude to the ITC Foundation Special Scholarship (ITCFSS), which provided me a grant from the beginning to the end of my MSc study at the University of Twente, Faculty of Geo-Information Science and Earth Observation. I also would like to thank the Faculty of Geo-information Science and Earth Observation Science (ITC) Department of Natural Resources Management staff and students for all the supports and happy time we spent together at ITC Enschede, The Netherlands.

Lastly, I am also highly grateful to my families and friends for their love, prayers, and guidance during my entire MSc study.

TABLE OF CONTENTS

1. INTRODUCTION 1
1.1. Background 1
1.2. Literature review 3
1.3. Problem statement 5
1.4. Research objectives 8
1.4.1. General objective: 8
1.4.2. Specific objectives, research questions (RQ), and research hypothesis (Ho): 8
1.5. Conceptual diagram 9
2. MATERIALS AND METHODS 11
2.1. Study area and research data 11
2.1.1. Field data/in situ measurement 11
2.1.2. Sentinel-1 and Sentinel-2 data acquisition and pre-processing 11
2.2. Methods 13
2.2.1. Sentinel-1 and Sentinel-2 feature extraction and data analysis 14
2.2.2. Statistical data analysis 16
2.2.3. Random Forest classification of Sentinel-1, Sentinel-2, and their combination data 16
3. RESULT 18
3.1. Backscatter properties of healthy and lodged wheat at different growth stages 18
3.2. Spectral properties of healthy and lodged wheat at different growth stages 20
3.3. Effect of lodging on vegetation indices 21
3.4. Kruskal Wallis and post hoc test results of Sentinel-1 and Sentinel-2 data 21
3.5. Random Forest classification results 23
3.5.1. Random Forest LS prediction map 25
4. DISCUSSION 26
4.1. Backscatter and spectral behaviour of wheat 26
4.2. Random Forest classification 27
4.3. Did the combination of Sentinel-1 and Sentinel-2 data improve classification accuracy? 28
4.4. Limitations and future works 29
4.5. Implications of the study 29
5. CONCLUSION 31
REFERENCES 33
APPENDICES 37

LIST OF FIGURES

Figure 1:Types of lodging and major lodging inducing factors 2
Figure 2: Random Forest ensemble process 5
Figure 3: Conceptual diagram........ 9
Figure 4: Study area and boundaries of wheat sample plots overlaid on Sentinel-2 images 12
Figure 5: Methodological flowchart 14
Figure 6: Box plot presenting the variation of backscatter for healthy wheat at stem elongation and booting stages 18
Figure 7: Box plot presenting the variation of backscatter for healthy and lodged wheat at flowering and milking stages 19
Figure 8: Box plot showing the variation of backscatter of healthy and lodged wheat at ripening growth stage 19
Figure 9: Average spectral reflectance of healthy wheat at stem elongation and booting stages and healthy and lodged wheat at flowering stage 20
Figure 10: Average spectral reflectance of healthy and lodged wheat at milking and ripening stages 21
Figure 11: Sentinel-2 derived vegetation indices of healthy and lodged wheat samples from stem elongation to ripening growth stages 21
Figure 12: Bonifiche Ferraresi farm wheat LS map predicted from the combination of Sentinel-1 and Sentinel-2 data on 19 May 2018 25

LIST OF TABLES

Table 1: RS based crop lodging studies in 2019 and 2020 6
Table 2: Sentinel-1 and Sentinel-2 image acquisition dates 13
Table 3: Specifications of Sentinel-1 and Sentinel-2 images 13
Table 4: Sentinel-1 and Sentinel-2 features 14
Table 5: No of Sentinel-1 and Sentinel-2 samples 15
Table 6: List of software used in this MSc research 15
Table 7: Vegetation indices 15
Table 8: Mean backscatter of wheat at different growth stage 19
Table 9: Average spectral reflectances of wheat at different wheat growth stages 20
Table 10: Kruskal Wallis test result for Sentinel-1 data 22
Table 11: Post hoc test result of Sentinel-1 data 22
Table 12: Kruskal Wallis test result for Sentinel-2 data 22
Table 13: Post hoc test result of Sentinel-2 data 23
Table 14: Confusion matrix and OOB error rate of Random Forest classifier 23
Table 15: Variable of importance for the Random Forest Classifier 24
Table 16: Random Forest classifier prediction outputs 24

1. INTRODUCTION

1.1. Background

Crop lodging refers to the permanent dislocation of the crop stem or root from its original position (Pinthus, 1974), and it is one of the yield-limiting factors in cereal crops like wheat, barley, and corn (Berry & Spink, 2012). It may occur at different growth stages of the crop (Berry et al., 2003); for example, in wheat crops, the event can occur at the end of the booting stage/beginning of the flowering stage (Chauhan, et al., 2020c). Lodging can be caused by three major factors, genetic factors, environmental factors, and poor crop management practices. As depicted in Figure 1, environmental factors include strong winds, heavy rain, very wet soil during late grain filling, and root or stem rots that weaken the plant base and cause severe yield loss (Rawson & Macpherson, 2000). On the other hand, genetic and poor management-related factors include planting tall thin stemmed varieties, excess nitrogen fertilizer application, dense planting, and sub-optimal planting time, which can increase the incidence of lodging (Pinthus, 1974). Although each factor has its impact, the combination of different factors increases the susceptibility of an area for an intense lodging event. For instance, the combination of strong winds, excess water content, high plant population density, excessive soil nitrogen content, and increased crop height can result in severe lodging events (Pinthus, 1974; Niu et al., 2016; Xiang et al., 2016). As a result, it would cause a reduction of profitability through reduced yield, delayed harvest, increased grain drying costs, and reduced grain quality (Berry & Spink, 2012; Yang et al., 2015). A study by Berry & Spink, (2012) indicated lodging induced yield loss in a wheat crop could reach 60-80%, and the total loss has been estimated at $80M per year in the UK (Berry & Spink, 2012). To mitigate this problem, several conventional lodging detection techniques have been developed.

Conventional lodging detection includes field-based visual inspection and laboratory analysis. Mostly, agronomists and plant physiologists practiced these field-based visual inspections and laboratory analysis to detect the root causes of the phenomenon. However, these field and lab-based techniques have several limitations; some limitations include limited spatial coverage, bias in visual rating, and poor accessibility in dangerous areas (Chauhan et al., 2019a). Therefore, to avoid these constraints and ensure sustainable agricultural production of the crop, selection, and application of suitable crop lodging assessment criteria are needed (Chauhan, et al., 2020c). According to a review by Chauhan et al., (2019a) one of the best complementary approaches is the application of remote sensing (RS) in lodging detection. The application of RS in wheat lodging detection can provide farmers with information about the crop status through the delivery of precise and continuous spatiotemporal data. As a result, it can help farmers to monitor and mitigate its knock-on effects such as destruction of plant morphology, physiological disruptions, and deterioration of grain quality. Henceforth, it would help increase the farmers' productivity (Fang & Cao, 2014). In addition, it can be used as proof for insurance claim adjustments and to reduce disagreements between farmers and insurance companies (Li et al., 2014; Vescovo et al., 2016).

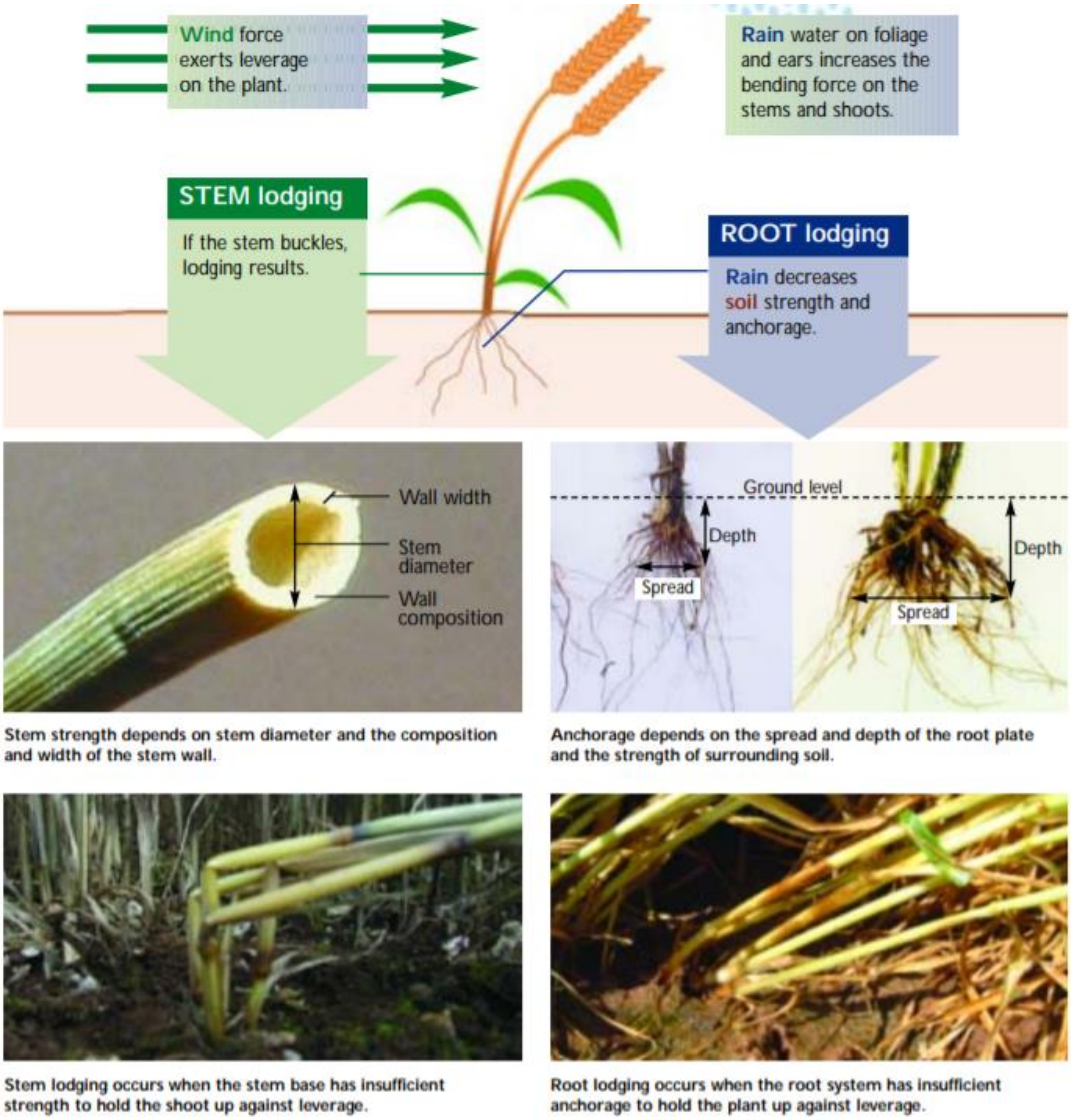


Figure 1:Types of lodging and major lodging inducing factors

Source: AHDB, (2005)

1.2. Literature review

RS is an increasingly important component of agricultural monitoring (Atzberger, 2013), and it can be used for lodging detection (Bouman, 1991). According to a review by Chauhan et al., (2019a) the RS platforms involved in lodging detection can be classified into three broad categories. These categories consisted of ground-based, airborne, and spaceborne platforms. Some of the RS-based crop lodging studies that were done after Chauhan et al., (2019a) review i.e., from 2019 to 2020, are presented in Table 1. Besides, various methods that were used in the classification of the RS images are presented below.

Ground-based RS platforms include hand-held sensors that can retrieve the target object information within a few centimetre distances. They are used to examine the effects of specific crop parameters by manipulating ground conditions, reducing the mixed-pixel effect, and by revisiting at any time (Moran, Inoue, and Barnes, 1997). Several studies have used ground-based RS platforms to explore the effect of lodging. For example, Fitch et al., (1984) studied the linear polarization of reflected light from common wheat, barley, and durum wheat in response to crop structural change. The study was done at the experimental research field level by using Kodak Plus-X Panchromatic film. The result of this study showed a reduced linear polarization for barley and an increased polarization for both kinds of wheat crops due to lodging. In addition, Bouman & van Kasteren, (1990) studied the influence of lodging on wheat, barley, and oats backscatter responses by using different incidence angles of a ground-based X-band, VV (vertical, vertical), and HH (horizontal, horizontal) polarization radar dataset. The study was done at a local scale in three different sites, Wageningen, Randwijk, and Dronten. The result showed a decrease in σVV° and σHH° of wheat and barley until it fluctuated from grain filling to the canopy was dying. On the other hand, the σVV°of oats decreased initially at its vegetative growth stage and suddenly increased to a stable level with the panicles' appearance. Although a ground-based or proximal sensing system closely monitors specific features like lodging without a mixed pixel effect, it has a limitation in terms of broad spatial coverage. Therefore, a combination of ground sensors with sensors covering large areas can be beneficial (Constantinescu et al., 2017).

Airborne RS includes airborne video imaging systems, light detection and ranging (LiDAR) /Radio Detection And Ranging (RADAR) that can be used for monitoring different natural and man-made phenomenon (Chauhan et al., 2019a). However, the recent development of Unmanned Aerial Vehicle (UAV) sensors in the airborne RS sector provides new opportunities to monitor agricultural production constraints (Colomina et al., 2008). UAVs provide high spatial and temporal resolution data that can be used for monitoring of lodging. Several lodging studies have been performed using airborne UAV data. For example, Liu et al., (2014) studied the potential of combining the spectral and textural features of UAV images in the extraction of the lodged wheat area. They found that combining the two features can clearly differentiate lodged and non-lodged wheat with the highest classification accuracy. Besides, Chauhan et al., (2019b) demonstrated the potential of multispectral UAV data in wheat lodging detection by analysing different lodging severity grades by using the nearest neighbourhood classification algorithm. The result showed an increase in the magnitude of reflectance in red-edge and Near-infrared (NIR) bands due to lodging severity. UAVs provide numerous advantages, especially in precision farming; however, they are also associated with some problems. For instance, the instrument’s lightweight makes it very prone to harsh weather conditions like strong wind and rainfall; thus, a data gap may occur. Besides, UAVs can collect data over a few hectares. Therefore, to cover large areas and reduce the effect of harsh weather conditions, the combination of UAV data with other sensors may be required.

Spaceborne platforms offer spatial coverage over larger areas and regular temporal coverage (Chauhan et al., 2019a). Different spaceborne sensors have been involved in land monitoring, including agriculture; however, the launch of the European Space Agency (ESA) Copernicus Sentinel-1 and Sentinel-2 missions offer a new opportunity to investigate agricultural production bottlenecks (Malenovský et al., 2012). Sentinel-1 is a Synthetic Aperture Radar (SAR) data. It has the advantage of operating day and night under all weather conditions to help in real-time monitoring of different phenomena, including lodging (Torres et al., 2012). Few lodging-related studies have been conducted by combinations of Sentinel-1 and other spaceborne sensors data. For instance, Chauhan, et al., (2020a) used multisensor SAR data (Sentinel-1 and RADARSAT-2) with a support vector regression (SVR) to estimate wheat crop angle of inclination as an indicator of wheat lodging. The authors found Sentinel-1 has comparable accuracy with RADARSAT-2 FQ21. Besides, Chauhan et al., (2020b) studied the potential of Sentinel-1 data and RADARSAT-2 data with partial least square discriminative analysis (PLS-DA) in wheat lodging severity classification. They found comparable overall accuracy between Sentinel-1 and RADARSAT-2 FQ21.

Sentinel-2 is another European Space Agency (ESA) Copernicus mission, which provides high-resolution multispectral (13 spectral bands) optical imagery to detect different phenomena on earth. Besides, Sentinel-2 has a high revisit time that can be used for real-time monitoring of various phenomena like lodging (Drusch et al., 2012). For example, Chauhan et al., (2020c) used Sentinel-2 and Sentinel-1 time-series data and statistical analysis to understand the influence of lodging on spectral reflectance and backscatter/coherence of Sentinel-2 and Sentinel-1, respectively (Table 1). Both Sentinel-1 and Sentinel-2 data performed very well in the classification of the wheat lodging classes. However, no study investigated the potential Sentinel-1, Sentinel-2, and their combination data in wheat lodging detection and classification using the Random Forest classifier (Shu et al., 2020).

Random Forest classifier is one of the most popular machine learning algorithms in classifying different RS images (Horning, 2010). It is a supervised classification algorithm that compiles/ensembles all decision tree outputs in the input vector classification (Figure 2). The decision trees are predictive models trained by drawing of random variables from the training samples. Each tree casts a unit vote for the most popular class to classify an input vector (Breiman, 2001).

Random Forest provides various benefits compared to decision trees, maximum likelihood classifier, and other machine learning algorithms. Random Forest can: 1) give very high accuracy, 2) it is not sensitive to overfitting of data, 3) it can handle categorical data and missing values, 4) it does not need pre-processing, 5) it is suitable with large datasets, 6) it can handle multicollinearity problems, and 7) it has high processing speed (Belgiu & Drăgu, 2016; Breiman, 2001; Gislason et al., 2006; Pal, 2005; Horning, 2010). Zhang et al. (2020) used UAV imagery and different machine learning algorithms (Random Forest, neural network, and support vector machine) to detect wheat lodging. The authors found comparable accuracy between Random Forest and neural network (GoogLeNet) machine learning algorithms. Besides, Phillips & Ward, (2020) compared Random Forest and Artificial Neural Networks (ANN) machine learning algorithms to detect maize crop lodging using UAV images. They found that the Random Forest model performed best as compared to ANN. Furthermore, Zhou et al., (2020) investigated the potential of multitemporal Gaofen-1 (GF-1) optical satellite images with Random Forest and partial least square modelling techniques for regional-scale maize lodging modelling and monitoring. The result showed that the Random Forest model performed better than the partial least square model.

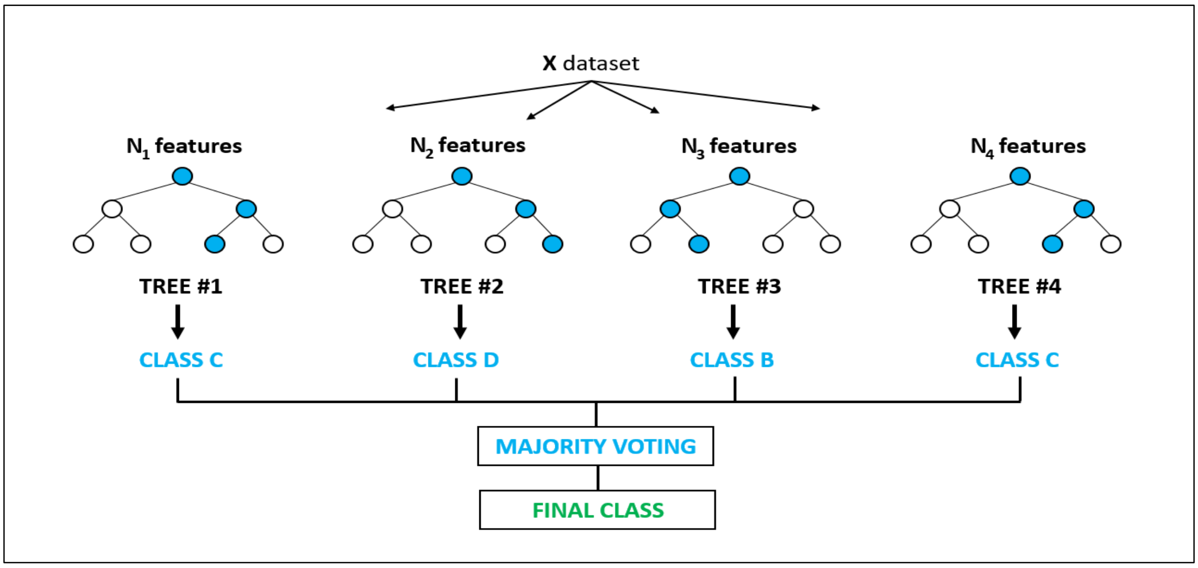


Figure 2: Random Forest ensemble process

1.3. Problem statement

Based on the above literature, Random Forest is a suitable algorithm in lodging detection and classification (Phillips & Ward, 2020; Zhou et al., 2020). However, its potential applications in the classification of spaceborne Sentinel-1, Sentinel-2, and their combination data into different wheat lodging classes based on LS were not studied. Therefore, the aim of the study was to evaluate the potential of Sentinel-1, Sentinel-2, and their combination data in wheat lodging detection and classification using a Random Forest classifier.

Table 1: RS based crop lodging studies in 2019 and 2020

Platform	Sensor	Pixel Size and No of Bands	No of Observation	Crop	Study Area	Size	Aim	Result	Author/s
Airborne and Spaceborne	Low Orbiting Satellite (LOS) and Small UAV-Based High-Resolution Imagery Data	UAS GSD= 0.01 m and 0.03 m and with 12.4 mega-pixel resolution in the VIS spectral band LOS GSD= 3.00 m with VIS and NIR spectral band	Two	Irrigated Spearmint	Toppenish, Washington, USA	Small field	To evaluate the performance of small UAV and LOS imagery in quantifying lodging in irrigated spearmint crop by using height and colour features from small UAS imagery and colour features from LOS, respectively	CSM derived from small UAV-based imagery performed better than the image colour features for spearmint lodging assessments. LOS can also be used for large area crop lodging assessment	Vargas et al., (2020)
Spaceborne	RADARSAT-2 and Sentinel-1	RADARSAT-2= 7×7 m with C-band Sentinel-1=10×10m with C- band	Five RADARSAT-2 images and Eleven Sentinel-1 images were acquired over the study area between30 May 2018 and 30 Jun 2018	-Durum wheat, -Soft wheat	Ferrara, Italy	3850 ha	To compare the performance of Sentinel-1 and multi-incidence angle RADARSAT-2 data for estimating CAI as an indicator of wheat lodging detection	RADARSAT-2 FQ8 data perform a great prediction of CAI than RADARSAT-2 FQ21 and Sentinel-1 data. However, Sentinel-1 and RADARSAT-2 FQ21 showed a comparable result	Chauhan, et al., (2020a)
Spaceborne	Sentinel-1 and Sentinel-2	Sentinel-1=15×15m with C-band Sentinel-2=10×10m With Thirteen spectral bands in the VIS, red edge, NIR, and SWIR domains	Nineteen (19) Sentinel-1 and Eight (8) Sentinel-2 images were acquired over the study area between 14 Mar 2018 and 30 Jun 2018	- Durum wheat, -Soft wheat	Ferrara, Italy	3850 ha	To understand lodging induced changes on the backscatter and spectral reflectance of Sentinel-1 and Sentinel-2, respectively	Sentinel-1 can best discriminate lodged and non-lodged classes as compared to Sentinel-2	Chauhan et al., (2020c)

Spaceborne	Gaofen-1 (GF-1)	16×16 m with four Spectral bands including Blue, Green, Red, and NIR	Two images were acquired on 23 Aug 2018 before lodging and 8 Sept 2018 after lodging, respectively	Maize	Gaocheng, Shijiazhuang City, Hebei Province, China	549 km²	To identify the potential of Gaofen-1 (GF-1) optical imagery with Random Forest and Partial Least Square models in maize lodging modelling and monitoring	The Random Forest model performed better than the partial least square model in classification and modelling of the acquired image	Zhou et al., (2020)
Spaceborne	Sentinel-1	10×10m With C-band	Two Sentinel-1 images were acquired on 31 Aug 2018 before lodging and 12 Sept 2018, respectively	Maize	Goacheng District, Shijiazhuang City, Hebei Province in North China	549 km²	To calculate maize lodging angle and monitor maize lodging event based on the inversion of plant height results before and after lodging from dual-polarization S-1 data	The ratio of VH and VV and also VV polarization of Sentinel-1 detect the change of maize height before and after maize lodging	Shu et al., (2020)

1.4. Research objectives

1.4.1. General objective:

The general objective of this study was to evaluate the potential of Sentinel-1, Sentinel-2, and their combination data in wheat lodging detection and classification by using a Random Forest classifier.

1.4.2. Specific objectives, research questions (RQ), and research hypothesis (Ho):

Objective 1: To evaluate the performance of Sentinel-1 data in wheat lodging detection and classification using Random Forest classifier

> **RQ 1:** Which Sentinel-1 parameter (e.g., σVV°, σVH°, σVH/VV°) best distinguishes the lodged wheat area from the non-lodged wheat area using Random Forest classifier?
>
> **Ho 1:** There is no significant difference between parameters derived from Sentinel-1 data for distinguishing the lodged wheat area from the non-lodged wheat area using Random Forest classifier

Objective 2: To evaluate the performance of Sentinel-2 data in wheat lodging detection and classification using Random Forest classifier

> **RQ 2:** Which Sentinel-2 parameter (e.g., the reflectance of spectral bands) best distinguishes the lodged wheat area from the non-lodged wheat area using Random Forest classifier?
>
> **Ho 2:** There is no significant difference between parameters derived from Sentinel-2 data for distinguishing the lodged wheat area from the non-lodged wheat area using Random Forest classifier

Objective 3: To compare the performance of Sentinel-1 and Sentinel-2 data combination for wheat lodging detection and classification using Random Forest classifier versus Sentinel- and Sentinel-2 data alone

> **RQ 3:** How is the accuracy (e.g., overall accuracy, kappa coefficient) of Sentinel-2 data compared to Sentinel-1 data for wheat lodging detection and classification?
>
> **Ho 3:** There is no significant difference between the accuracy (e.g., overall accuracy, kappa coefficient) of Sentinel-2 and Sentinel-1 data for wheat lodging detection and classification
>
> **RQ 4:** Will the combination of Sentinel-1 and Sentinel-2 data improve the wheat lodging detection and classification accuracy (e.g., overall accuracy, kappa coefficient) as compared to Sentinel-1 and Sentinel-2 data only?
>
> **Ho 4:** There is no significant difference in the accuracy (e.g., overall accuracy, kappa coefficient) of wheat lodging detection and classification if we combine Sentinel-1 and Sentinel 2 data as compared to using Sentinel-1 or Sentinel-2 only

1.5. Conceptual diagram

The system boundary is the Bonifiche Ferraresi farm, located in Jolanda di Savoia Ferrara, Italy. The system consists of two components: internal (farmers and arable land) and external (weather, RS platforms, RS scientists, insurance companies) components. Each component interacts with each other; for instance, as indicated in Figure 3, wheat lodging is a phenomenon caused by the influence of extreme weather events like a combination of strong wind and high rainfall. Therefore, for real-time detection of the event, the application of high spatial and temporal resolution RS data is vital. Besides, the active involvement of RS scientists in processing and analysing of these RS data is required for providing real-time information on the wheat lodging event. Moreover, farmers need to be active in using real-time RS information and taking accurate measures against lodging event. This study specifically investigated the potential of Sentinel-1, Sentinel-2, and their combination data in wheat lodging detection and classification by using the Random Forest classifier.

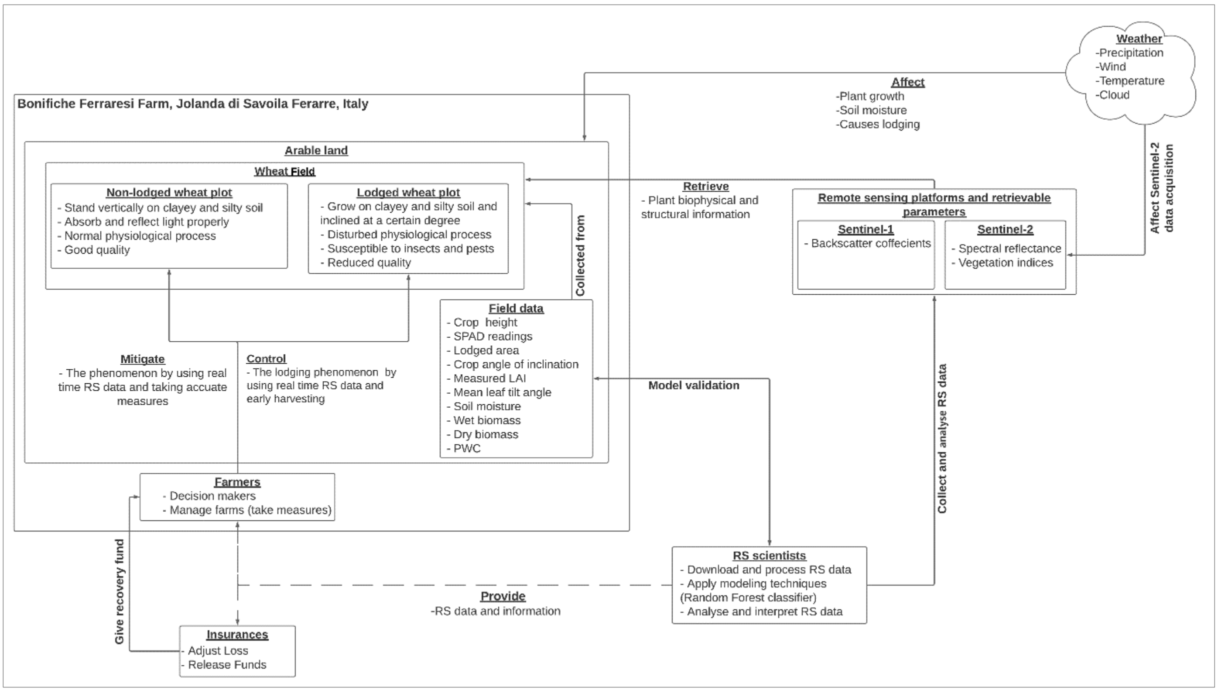


Figure 3: Conceptual diagram

2. MATERIALS AND METHODS

This chapter presents the data and methods used in this research. The first section presents the study area, field data, and remote sensing data. The second section presents the statistical data analysis and the procedures followed for Random Forest classification of lodged and healthy wheat samples using Sentinel-1, Sentinel-2, and their combination data.

2.1. Study area and research data

The study area is the Bonifiche Ferraresi farm (Figure 4), located in Jolanda di Savoia Ferrara, Italy, with a central coordinate of 44°52'59"N, 11°58'48"E. The area is characterized by its warm and temperate climate. The mean annual temperature and precipitation of the area are 13.6 °C and 691mm, respectively. The size of the arable land is 3850 ha, and it consists of two major soil types: clayey and silty soils. The farm is cultivated with different crops, including wheat, corn, rice, barley, soybean, potatoes, legumes, and other horticultural crops, but this study focused on winter wheat varieties.

2.1.1. Field data/in situ measurement

The field data were collected by Dr. Sugandh Chauhan, a previous PhD student at ITC. Seventy-six (76) winter wheat plots with a size of 60×60m were selected through stratified random sampling in her study. The plots were planted with Altamira, Bologna, Claudio, Giorgione, Marco Aurelio, Massimo Meridio, Monastir, Odisseo, PR22D66, Rebelde, and Senatore Capelli winter wheat varieties. The field data acquisition was performed from 14 Mar 2018 to the end of June 2018, including five growing stages (i.e., stem elongation, booting, flowering, milking, and ripening) of winter wheat. The measured field data included crop angle of inclination (CAI) and lodged area (LA) in percentage. According to Chauhan et al., (2020a) "CAI is defined as the angle made by the crop stem with respect to the vertical". These two field parameters were used to calculate lodging score (LS) (equation 1). The LS was used to define the lodging categories or groups and it was modified after Chauhan et al., (2020c) work. As such, healthy (He) wheat plots include plots that have an LS value equal to 0.0 (LS=0.0), moderately lodged (ML) wheat plots include plots that have an LS value of less than or equal to 0.30 ($0.0 < LS \leq 0.30$), a severely lodged (SL) wheat plots include plots that have an LS value of between 0.30 and 1 ($0.31 < LS \leq 1$).

$$\boldsymbol{LS = \frac{LA}{100} \times \frac{CAI}{90°}} \quad \textbf{(1)}$$

2.1.2. Sentinel-1 and Sentinel-2 data acquisition and pre-processing

Nineteen Sentinel-1 (A/B) and five Sentinel-2 (A/B) images were acquired between 14 Mar 2018 and 30 Jun 2018 (Table 2) from the ESA Copernicus Open Access Hub by Dr. Sugandh Chauhan. The Sentinel-1 images were acquired in ascending pass (ASC) and Interferometric Wide Swath (IW) mode with dual-polarization (VV, VH) (Table 3). Ground Range Detected (GRD) format Sentinel-1 data were acquired to extract the backscatter coefficient. The procedures outlined in Nelson et al., (2014) were applied in SARscape 5.5 software to pre-process the acquired GRD format Sentinel-1 data. The Sentinel-2 multispectral images were level 2A products (Bottom of atmosphere (BOA) reflectance data) (Table 3). Three spectral bands of Sentinel-2 data, namely band 1 (B1), band 9 (B9), and band 10 (B10), were removed since they were not important for this study.

The acquisition of Sentinel-1 and Sentinel-2 images were temporally aligned with field observation dates in two different ways: if the samples were lodged, Sentinel-1 and Sentinel-2 images acquired on the same date of field measurement or after the field measurement date was selected, if the samples were not lodged, Sentinel-1 and Sentinel-2 images acquired either on the same date of field measurement or before the field measurement date was selected. However, we only used five cloud-free Sentinel-2 images as compared to the 19 Sentinel-1 images.

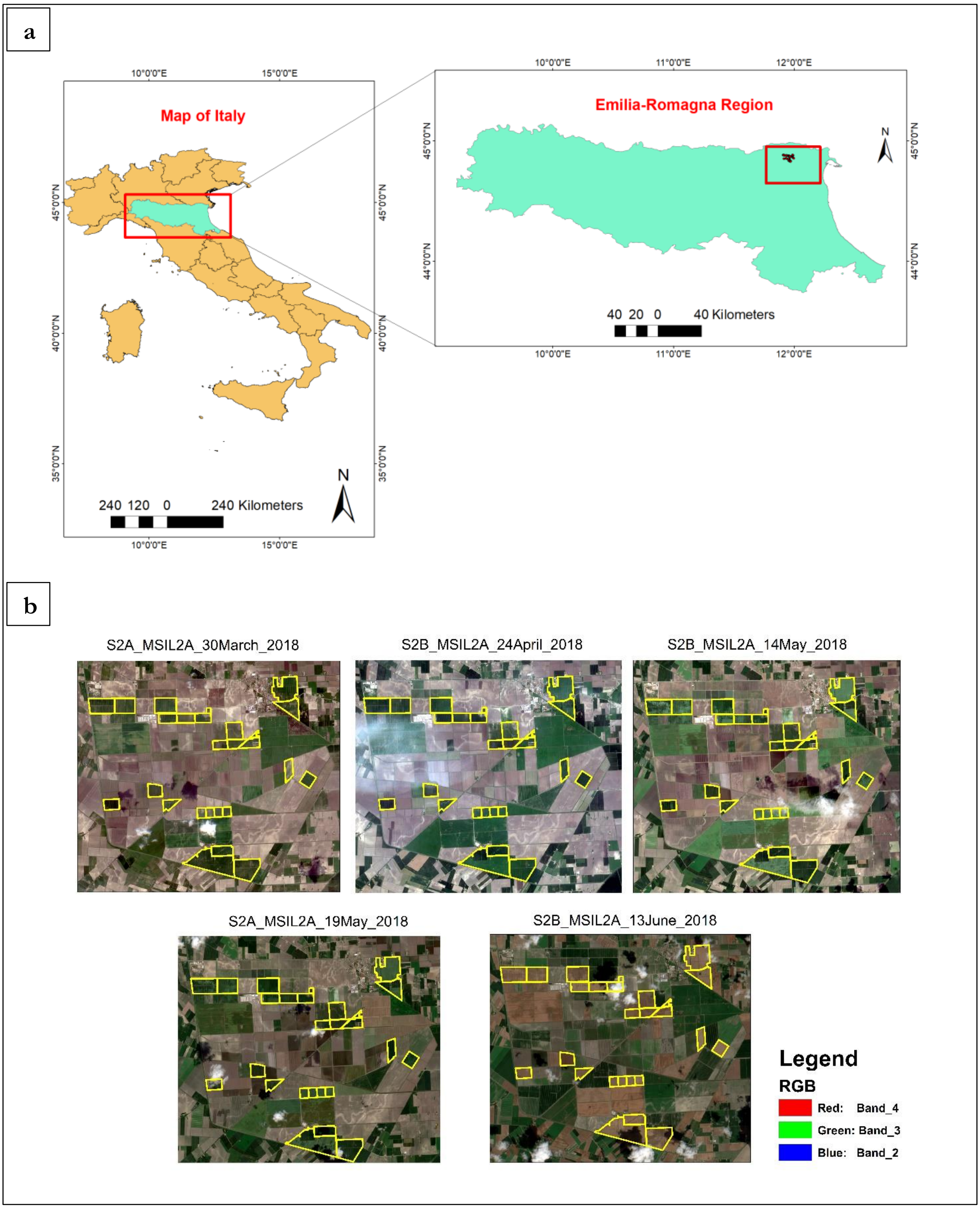


Figure 4: Study area (Bonifiche Ferraresi farm) (a), and (b) boundaries of wheat sample plots (in yellow) overlaid on Sentinel-2 images acquired at different wheat-growing stages

Table 2: Sentinel-1 and Sentinel-2 image acquisition dates

	Sentinel-1				Sentinel-2			
Date	**March**	**April**	**May**	**June**	**March**	**April**	**May**	**June**
1		✓	✓					
6				✓				
7		✓	✓					
12				✓				
13		✓	✓					✓
14	✓						✓	
18				✓				
19		✓	✓				✓	
20	✓							
24				✓		✓		
25		✓	✓					
26	✓							
30				✓	✓			
31			✓					

Table 3: Specifications of Sentinel-1 and Sentinel-2 images

Sentinel-1		Sentinel-2			
Parameter	**Specification**	**Spectral band**	**central wavelength (nm)**	**Bandwidth (nm)**	**Spatial resolution (m)**
Wavelength	C-band	B1 Coastal aerosol	443	20	60
Frequency	5.405 GHz	B2 Blue	490	65	10
Product type	GRD, SLC	B3 Green	560	35	10
Acquisition mode	IW	B4 Red	665	30	10
Incidence angle	39.7–40.4°	B5 Red edge1 (RE-1)	705	15	20
Pass	ASC	B6 Red edge2 (RE-2)	740	15	20
Polarisation	VH, VV	B7 Red edge3 (RE-3)	783	20	20
Spatial resolution (resampled)	15 m	B8 NIR-1	842	115	10
Repeat cycle	6 days	B8a NIR-2	865	20	20
		B9 Water vapor	940	20	60
		B10 SWIR Cirrus	1375	30	60
		B11 SWIR-1	1610	90	20
		B12 SWIR-2	2190	180	20

2.2. Methods

The Random Forest machine learning algorithm was used to predict LS in wheat fields based on the relationship between RS data (Sentinel-1, Sentinel-2, and their combination) and observed lodging in wheat plots. The different steps performed to predict the LS include feature extraction, feature splitting, optimal input predictor variables selection, the Random Forest model training, and finally, Random Forest model validation (Figure 5).

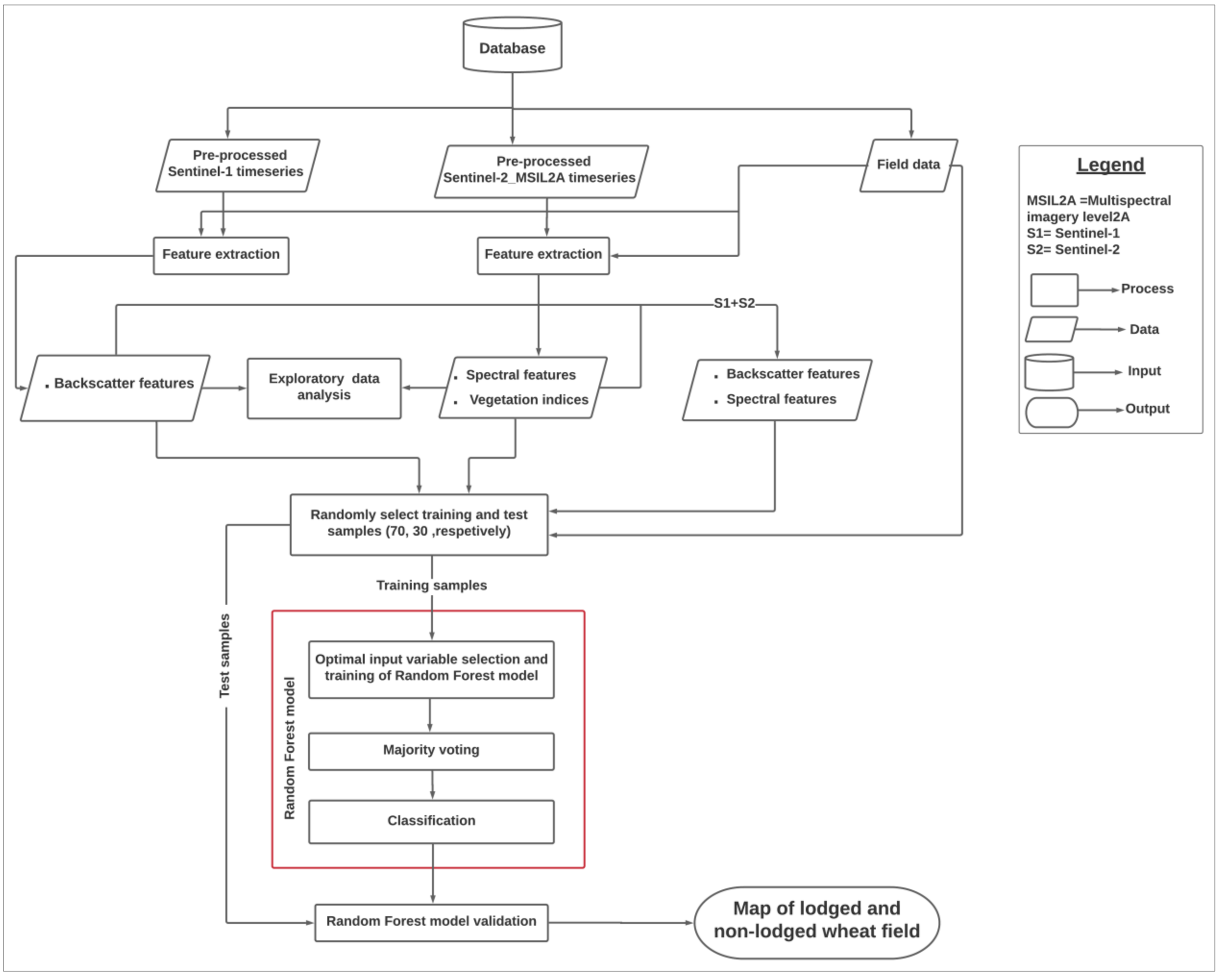


Figure 5: Methodological flowchart

2.2.1. Sentinel-1 and Sentinel-2 feature extraction and data analysis

The mean backscatter signature or features (Table 4) of nineteen Sentinel-1 images were extracted from the locations where wheat field samples were taken with a 3×3 kernel (window) size. The kernel size was based on the size of the field sample plots (60m×60m). The spectral extraction tool developed at the NRS department in ENVI Classic version 5.5.3 software (Table 6) was used to extract the backscatter signature of healthy (He) and lodged (ML and SL) wheat. Then the Sentinel-1 ratio (σVH/VV°) was calculated by subtracting the σVV° backscatter from σVH° (σVH°- σVV°). Then, the backscatter signature of the healthy and lodged wheat samples was explored by using box plots in Microsoft Excel 2020 (Table 6). The backscatter exploration was performed for five wheat growth stages (stem elongation, booting, flowering, milking, and ripening). However, since the first instance of lodging occurred at the beginning of the flowering stage, the lodged classes in flowering, milking, and ripening stages grouped into ML and SL based on the field-based lodging score (LS). In total, 228 Sentinel-1 samples were used for further analysis (Table 5).

Table 4: Sentinel-1 and Sentinel-2 features

Sentinel-1	Sentinel-2
VV, VH, VH/VV	Blue, Green, Red, RE-1, RE-2, RE-3, NIR-1, NIR-2, SWIR-1, SWIR-2, NDVI-1, NDVI-2, NDVIRE-1, NDVIRE-2, NDVIRE-3, NDWI, EVI, SAVI, DVI

Table 5: No of Sentinel-1 and Sentinel-2 samples

Lodging Class	Sentinel-1	Sentinel-2
	Sample No	Sample No
He	160	59
ML	13	13
SL	55	48
Total	228	120

The mean spectral reflectance or features (Table 4) of five Sentinel-2 images were extracted from the locations where wheat field samples were taken with a 5×5 kernel (window) size using the spectral extraction tool developed at the NRS department in the ENVI Classic version 5.5.3 software. The kernel size was based on the field sample plots size (60m×60m) and the Sentinel-2 spatial resolution (10m). Then the spectral analysis of wheat was performed for the same five growth stages using MATLAB 2019a and Microsoft Excel 2020 software (Table 6). Furthermore, several Sentinel-2 derived vegetation indices; NDVI-1, NDVI-2, NDVIRE-1, NDVIRE-2, NDVIRE-3, NDWI, RVI, EVI, SAVI (Table 7) were calculated, to explore their relationship with wheat lodging. In total, 120 Sentinel-2 samples were used for further analysis (Table 5).

Table 6: List of software used in this MSc research

Software	Function
SNAP 5.0	Sentinel-2 data Pre-processing
ENVI Classic 5.5.3	Sentinel-1 and Sentinel-2 features extraction
Microsoft Excel 2020	Data Analysis
MATLAB 2019a	Data Analysis
R version 4.0.3	Classification and Data Analysis
IBM SPSS Statistics 27	Statistical data analysis
Arc GIS 10.8	Visualization and Map Production

Table 7: Vegetation indices

Vegetation Index	Purpose	Equation	References
Normalized Difference Vegetation Index 1	Measure vegetation greenness and health	NDVI-1 = (NIR-1 - Red) / (NIR-1 + Red)	Rouse, (1974)
Normalized Difference Vegetation Index 2	Measure vegetation greenness and health	NDVI-2= (NIR-2 – Red) / (NIR-2 + Red)	Rouse, (1974)
Normalized Difference Vegetation Index red-edge 1 narrow	Measure vegetation healthiness and post fire assessment	NDVIRE-1 = (NIR-2 – RE-1) / (NIR-2 + RE-1)	Fernández-Manso et al., (2016)
Normalized Difference Vegetation Index red-edge 2 narrow	Measure vegetation healthiness and post fire assessment	NDVIRE-2 = (NIR-2 – RE-2) / (NIR-2 + RE-2)	Fernández-Manso et al., (2016)
Normalized Difference Vegetation Index red-edge 3 narrow	Measure vegetation healthiness and post fire assessment	NDVIRE-3= (NIR-2 – RE-3) / (NIR-2 + RE-3)	Fernández-Manso et al., (2016)

Normalized Difference Water Index	To estimate the plant moisture content	NDWI = (NIR-2 – SWIR-1) / (NIR-2 + SWIR-1)	Gao, (1996)
Ratio Vegetation Index	Green biomass estimations and monitoring	RVI= Red / NIR-2	Pearson & Miller, (1972)
Enhanced Vegetation Index	Great sensitivity to high biomass accumulations hence it is useful in areas where NDVI is saturated	EVI = ((NIR-1 – Red) × 2.5) / ((NIR-1 + 6 × Red – 7.5 × Blue) +1)	Huete et al., (1999)
Soil Adjusted vegetation Index	Minimize the effects of soil background on the vegetation signal	SAVI = ((NIR-1 - Red) × 1.5) / (NIR-1 + Red + 0.5)	Huete, (1988)
Difference Vegetation Index	Measure vegetation development	DVI = (NIR-1 – Red) × 2.4	Richardson & Wiegand, (1977)

2.2.2. Statistical data analysis

To explore the statistical relationship between the field-based LS and RS data (Sentinel-1 and Sentinel-2), a Kruskal Wallis and post hoc tests were performed in the IBM SPSS Statistics 27 (Table 6). The Kruskal Wallis test is a non-parametric statistical test that can be used for testing data that have a continuous distribution, but the data are ordinal and not interval (MacFarland & Yates, 2016). It can be used as a one-way ANOVA for data that does not have a normal distribution. To run the Kruskal Wallis and post hoc test for Sentinel-1 data, firstly, the field-based LS was assigned as the grouping variable and the Sentinel-1 backscatter data ($\sigma VV°$, $\sigma VH°$, and $\sigma VH°/\sigma VV°$) as testing variables. Then a Kruskal-Wallis one-way ANOVA (k samples) with multiple pairwise comparisons (post hoc) was used. The same procedures were repeated to test the Sentinel-2 data.

2.2.3. Random Forest classification of Sentinel-1, Sentinel-2, and their combination data

The Random Forest classifier was used to classify the Sentinel-1, Sentinel-2, and their combination data into healthy and two lodging classes (ML and SL). Prior to implementing the Random Forest classifier, the backscatter and spectral data of Sentinel-1, Sentinel-2, and their combination data, respectively, were saved in a comma-delimited (CSV) format in Microsoft Excel version 2020. Then, different packages were installed and loaded in R version 4.0.3 to run the Random Forest model. The installed packages include raster version 3.4-5 (Hijmans, 2020), randomForest version 4.6-14 (Wiener, 2002), sp version 1.4-5 (Pebesma & Bivand, 2005), rgdal version 1.5-23 (Rowlingson, 2021), ggplot2 version 3.3.3 (Wickham, 2016), and caret version 0.4-3 (Kuhn, 2020).

The other preliminary steps of the Random Forest classifier include feature splitting, differentiation of the dependent and explanatory variables, and input predictor variables selections.

First, the Sentinel-1, Sentinel-2, and their combination data were split into training and test samples in R. The splitting was done in a proportion of 70/ 30 i.e., 70% training samples and 30% test samples. Then "as.factor" function of R version 4.0.3 was used to convert the LS column or the explanatory variable of Sentinel-1, Sentinel-2, and their combination data, respectively, into a factor variable and to run a classification.

Then the optimal input predictor variables (mtry) were identified for the splitting of nodes in the Random Forest decision tree development by using Sentinel-1, Sentinel-2, and their combination data, respectively. A "tuneRF" function was used to find the optimal mtry by setting also other parameters ("stepFactor', "improve", "trace", "plot"). The "stepFactor" were used to define the inflated and deflated value of mtry at each iteration. The "improve" was used to define the OOB error improvement in search of better (small) OOB error. The "plot" setting was used to plot the OOB error as a function of "mtry" whereas, the "trace" option used to print the progress of the search. Secondly, the default ntree value of the Random Forests package (randomForest) in R version 4.0.3 (ntree=500) was used as an optimal ntree.

After the required parameters (mtry and ntree) were defined, the Random Forest model was trained by using Sentinel-1, Sentinel-2, and their combination data, respectively. Then the performance of the Random Forest model outputs of the three datasets separately was evaluated through their confusion matrix, out-of-bag error rate (OOB), and other parameters like variable importance. The variable of importance function was used to examine the importance of each Sentinel-1, Sentinel-2, and their combination data variables, respectively, in the Random Forest model (Breiman, 2001).

Lastly, the Random Forest models developed using Sentinel-1, Sentinel-2, and their combination data, respectively, were validated by using the 30% test samples of the three datasets separately. The predict function of R version 4.0.3 was used to run the Random Forest models prediction. The best performing validated model was used on 19 May 2018 Sentinel-1 and Sentinel-2 combination image of the study area to generate a wheat LS classification map. The main reason for selecting the 19 May 2018 Sentinel-1 and Sentinel-2 combination image was to have a common acquisition date and ease of interpretation.

3. RESULT

We first present the Sentinel-1 backscatter, Sentinel-2 spectral properties, and Sentinel-2 derived vegetation indices of healthy and lodged wheat at different wheat growth stages followed by the Kruskal Wallis and post hoc test results. The final section presents the Random Forest classifier training and prediction outputs for Sentinel-1, Sentinel-2, and their combination data.

3.1. Backscatter properties of healthy and lodged wheat at different growth stages

At the stem elongation phase (Figure 6a), the mean σVV° and σVH° of He wheat samples were -10.90 and -18.26, but when the wheat turns its phase to booting (Figure 6b), the mean σVV° and σVH° decreased to -13.69 and -19.66, respectively (Table 8). However, the σVH/VV° showed an increasing pattern from stem elongation to booting.

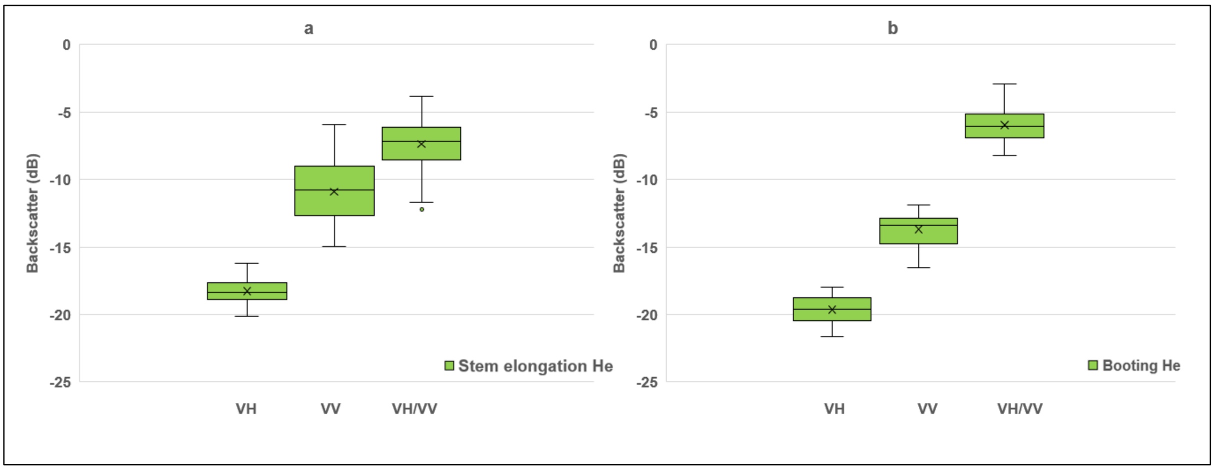


Figure 6: Box plot presenting the variation of backscatter (σVH°, σVV°, σVH/VV°) for healthy (He) wheat at (a) stem elongation (n=76) and (b) booting (n=33) stages

At the flowering stage (Figure 7a), the mean σVV° and σVH° of He wheat samples further decreased to -15.02 and -20.04, respectively due to structural parameter (e.g., leaf area index (LAI), fresh biomass (FB)) changes of wheat. But the σVH/VV° further increased to -5.01 (Table 8). On the other hand, the mean σVV°, σVH°, and σVH/VV° of SL wheat samples were higher than the mean σVV°, σVH°, and σVH/VV° of ML and He samples. Besides, the mean σVV° of ML wheat sample was higher than the He samples; however, the mean σVH° and σVH/VV° of ML wheat showed a decreasing pattern compared to He samples.

At the milking growth stage of wheat (Figure 7b), the mean σVV° and σVH° of He wheat samples increased to -13.00 and -18.44, respectively; however, the σVH/VV° decreased to -5.44. In addition, the mean σVV°, σVH°, and σVH/VV° of ML and SL samples increased to -10.83 and -9.86, -4.35, respectively. The rate of change of σVV°, σVH°, and σVH/VV° for ML and SL samples at milking wheat growth stage was higher than the flowering stage. The increment of the σVV°, σVH° was consistent with the lodging severity.

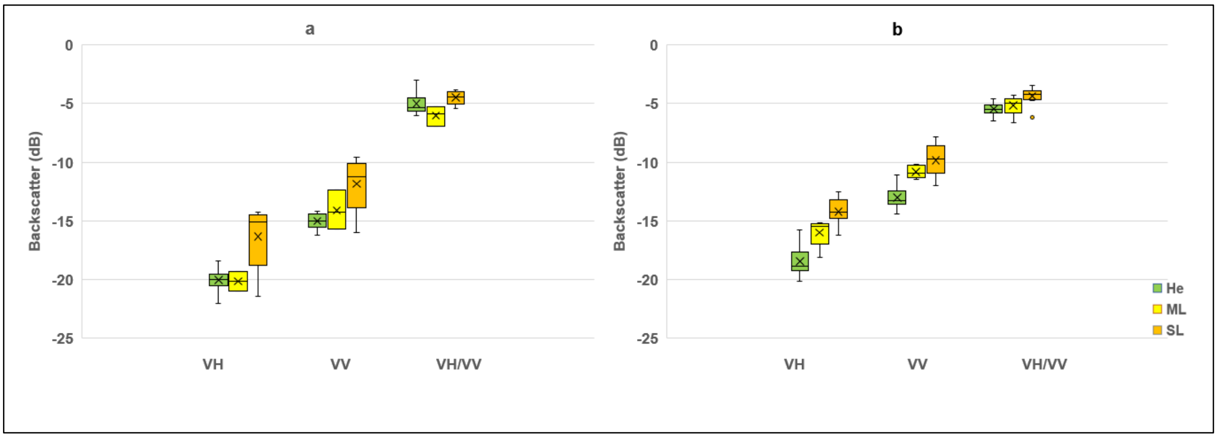


Figure 7: Box plot presenting the variation of backscatter (σVH°, σVV°, σVH/VV°) for healthy (He) and lodged (ML, SL) wheat at (a) flowering (He; n=17, ML; n=3, SL; n=5) and (b) milking (H; n=21, ML; n=5, SL; n=18) stages

Table 8: Mean backscatter of wheat at different growth stage

		Mean backscatter		
Growth stage	**LS class**	$\boldsymbol{\sigma VV°}$	$\boldsymbol{\sigma VH°}$	$\boldsymbol{\sigma VH/VV°}$
Stem elongation	He	-10.90	-18.26	-7.36
Booting	He	-13.69	-19.66	-5.96
Flowering	He	-15.02	-20.04	-5.01
	ML	-14.11	-20.14	-6.03
	SL	-11.84	-16.34	-4.50
Milking	He	-13.00	-18.44	-5.44
	ML	-10.83	-15.98	-5.15
	SL	-9.86	-14.21	-4.35
Ripening	He	-11.37	-18.85	-7.48
	ML	-11.45	-19.81	-8.36
	SL	-11.95	-19.71	-7.76

The mean σVV° of He wheat samples further increased at the ripening stage than the milking, flowering, and booting stages (Figure 8). However, the σVH° and σVH/VV° of He wheat samples decreased at the ripening stage compared to the milking stage of wheat. The mean σVV°, σVH°, and σVH/VV° of ML and SL wheat samples decreased compared to the milking stage.

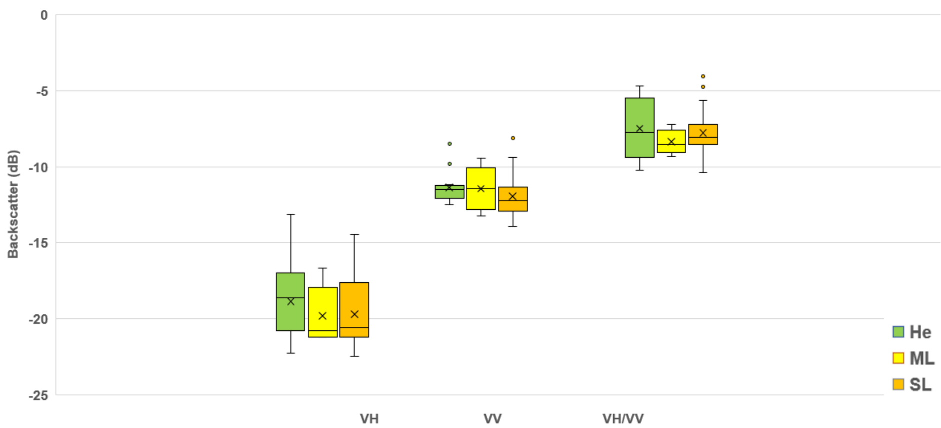


Figure 8: Box plot showing the variation of backscatter (σVH°, σVV°, σVH/VV°) of healthy (He, n=13) and lodged (ML; n=5, SL; n=32) wheat at ripening growth stage

3.2. Spectral properties of healthy and lodged wheat at different growth stages

The spectral reflectance of wheat can be affected by different factors such as the growth stages, pigment concentration, water content, and structural changes like lodging. The average spectral reflectance characteristics of He, ML, and SL wheat samples were analyzed at different growth stages to understand the effect of lodging on the Sentinel-2 spectral properties of wheat.

Since there were no lodged samples in the first two growing stages (stem elongation and booting) of wheat, the spectral analysis for those stages was only performed for He wheat samples. The average spectral reflectance of He wheat at the booting stage was higher than the stem elongation stage in all spectral regions (visible, RE, NIR, and SWIR) (Figure 9a). However, the rate of change in the NIR region was higher.

At the flowering (Figure 9b) and milking (Figure 10a) stages of wheat, the average spectral reflectance of the SL and ML wheat samples were higher than the He wheat samples in all spectral bands. However, the difference between the average spectral reflectance of He, ML, and SL wheat samples was easily differentiable at the milking stage of wheat compared to the flowering stage (Table 9).

Table 9: Average spectral reflectances of wheat at different wheat growth stages

		Sentinel-2 average spectral reflectance									
Growth-stage	**Class**	**Blue**	**Green**	**Red**	**RE-1**	**RE-2**	**RE-3**	**NIR-1**	**NIR-2**	**SWIR-1**	**SWIR-2**
Stem-elongation	**He**	0.05	0.07	0.05	0.11	0.25	0.30	0.31	0.31	0.14	0.08
Booting	**He**	0.06	0.08	0.06	0.10	0.28	0.39	0.39	0.40	0.15	0.08
Flowering	**He**	0.02	0.04	0.02	0.07	0.30	0.44	0.46	0.46	0.14	0.06
	ML	0.05	0.07	0.04	0.11	0.37	0.52	0.54	0.54	0.18	0.09
	SL	0.04	0.08	0.04	0.12	0.40	0.54	0.57	0.57	0.17	0.08
Milking	**He**	0.01	0.04	0.03	0.07	0.26	0.37	0.38	0.39	0.13	0.06
	ML	0.02	0.05	0.03	0.08	0.29	0.39	0.41	0.42	0.14	0.07
	SL	0.04	0.08	0.05	0.12	0.38	0.50	0.51	0.53	0.17	0.08
Ripening	**He**	0.05	0.08	0.12	0.15	0.16	0.19	0.20	0.22	0.24	0.16
	ML	0.07	0.11	0.16	0.19	0.20	0.23	0.24	0.26	0.28	0.19
	SL	0.09	0.13	0.18	0.23	0.25	0.28	0.30	0.32	0.32	0.22

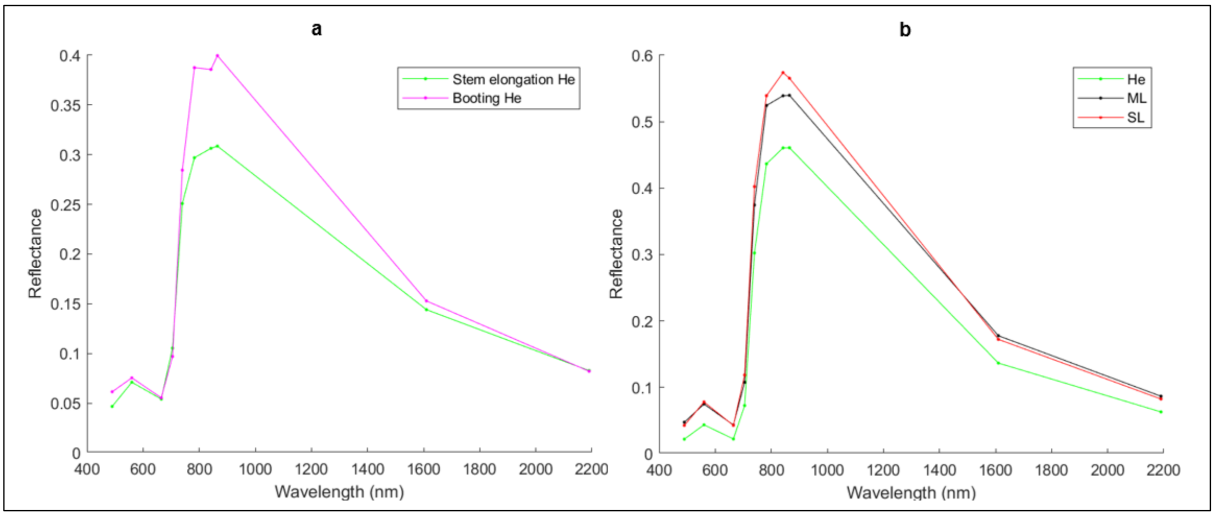


Figure 9: Average spectral reflectance of (a) healthy wheat at stem elongation and booting stages (He, n=15), (b) healthy (He, n=5), and lodged (ML; n=3, SL; n=5) wheat at flowering stage

Figure 10b shows the average spectral reflectance of He, ML, and SL wheat samples at the ripening growth stage. At this stage, the average spectral reflectance of wheat loses its normal spectral shape due to the loss of chlorophyll and water absorption peaks (Chauhan, et al., 2020c). Although the normal average spectral reflectance behaviour is changed, the average spectral reflectance difference between the He, ML, and SL wheat samples was noticeable in all spectral bands.

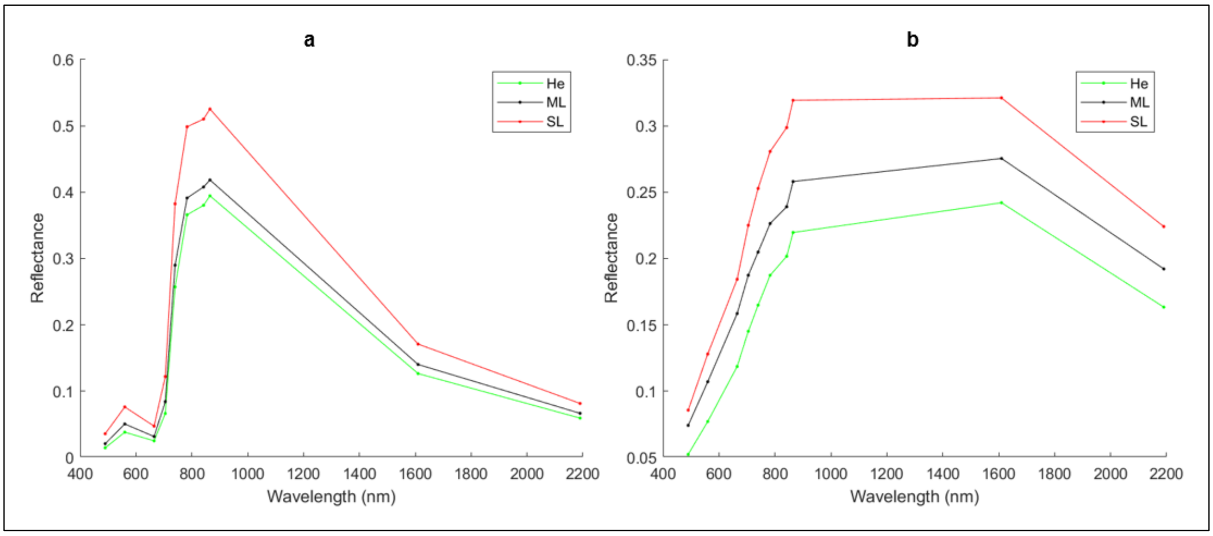


Figure 10: Average spectral reflectance of healthy (He) and lodged (ML, SL) wheat at (a) milking (He; n=15, ML; n=5, SL; n=16) and (b) ripening (H; n=9, ML; n=5, SL; n=27) stages

3.3. Effect of lodging on vegetation indices

We also explored the relationship between selected Sentinel-2 derived vegetation indices and wheat lodging. Figure 11 shows the variation of different vegetation indices in different lodging classes. The mean values of some indices (NDVI-1, NDVI-2, NDVIRE-1, NDVIRE-2, NDWI, EVI, SAVI, and DVI) decreased when the severity of lodging increased. On the other hand, some other indices such as NDVIRE-3, RVI increased.

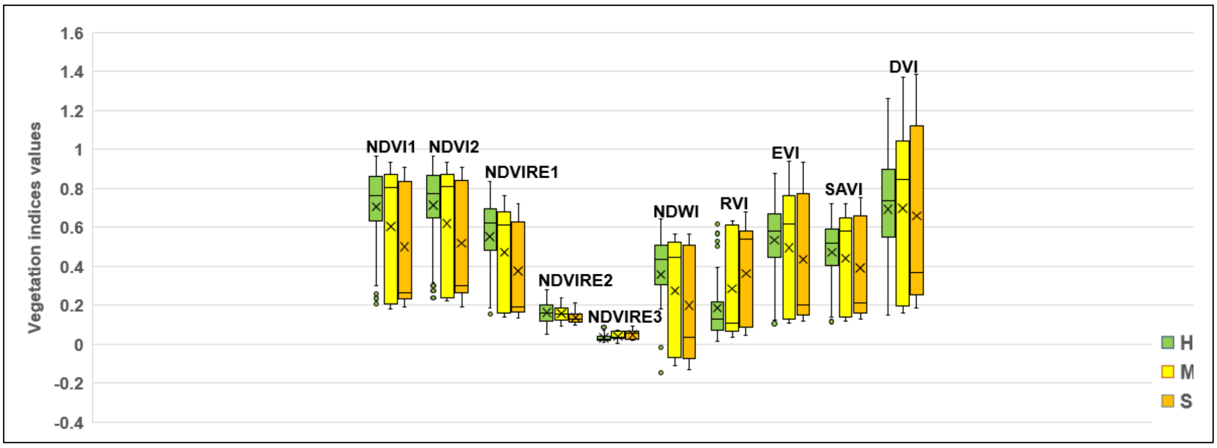


Figure 11: Sentinel-2 derived vegetation indices of healthy (He, n=59) and lodged (ML; n=13, SL; n=48) wheat samples from stem elongation to ripening growth stages

3.4. Kruskal Wallis and post hoc test results of Sentinel-1 and Sentinel-2 data

Table 10 presents the Kruskal Wallis test results of Sentinel-1 data. The σVV° was the only significant Sentinel-1 band at $p<0.05$; hence it could differentiate the He, ML, and SL wheat classes. However, the remaining two Sentinel-1 bands (σVH° and σVH/VV°) were not significant at $p<0.05$; hence it could not differentiate the three classes. On the other hand, the σVV°, σVH°, and σVH/σVV° showed different Kruskal Wallis test results in differentiating the He, ML, and SL wheat classes at different growth stages (flowering, milking, and ripening). For example, at the milking stage, all Sentinel-1 bands (σVV°, σVH°, and σVH/VV) were significant in differentiating the He, ML, and SL classes. However, at flowering and ripening stages, all Sentinel-1 bands (σVV°, σVH°, and σVH/VV°) were not significant.

The post hoc pairwise comparison result of all three Sentinel-1 bands presented in Table 11 reveals that σVV° could significantly differentiate the He from SL classes. Furthermore, the post hoc result at the milking growth stage showed, both σVV° and σVH° could differentiate He from ML and SL classes. In contrast, σVH/VV° could differentiate only the He from SL classes (Appendix Table 2).

Table 10: Kruskal Wallis test result for Sentinel-1 data

Lodging classes	P-value $\sigma VH°$	P-value $\sigma VV°$	P-value $\sigma VH/VV°$
H-ML-SL	0.14	0.00	0.61

Table 11: Post hoc test result of Sentinel-1 data

Class Pairs	P-value $\sigma VH°$	P-value $\sigma VV°$	P-value $\sigma VH/VV°$
H-ML	Na	.27	Na
H-SL	Na	.00	Na
ML-SL	Na	.46	Na

Table 12 shows the Kruskal Wallis test results of Sentinel-2 data. Most of the Sentinel-2 spectral bands were significant in terms of differentiating the three classes (He, ML, and SL). However, RE-3 and NIR-1 were not significant in terms of differentiating the three classes. Furthermore, the Kruskal Wallis test result of Sentinel-2 data at different wheat growth stages (flowering, milking, and ripening) revealed different results. Only the NIR-2 spectral band of Sentinel-2 was significant at the flowering stage to differentiate He, ML, and SL classes, while the other Sentinel-2 bands were non-significant. However, all Sentinel-2 spectral bands were significant in differentiating the three classes (Appendix Table 3) at milking and ripening stages.

The post hoc pairwise comparison result of the whole Sentinel-2 data revealed that most of the Sentinel-2 spectral bands (Green, Red, RE-1, SWIR-1, SWIR-2) could distinguish the He from SL classes and ML from the SL classes (Table 13). However, Blue, RE-2, and NIR-2 could only differentiate the He classes from SL classes. Since the two Sentinel-2 spectral bands were not significant in the Kruskal Wallis test, we could not find a post hoc pairwise comparison result; hence, we put "Na" meaning "missing values". Furthermore, the post hoc pairwise comparison result of Sentinel-2 data at the flowering stage showed NIR-2 could differentiate the He from SL classes, while the other spectral bands could not differentiate any class combination. However, all Sentinel-2 spectral bands could differentiate the He from SL classes at milking and ripening stages. Besides, most of the spectral bands could differentiate ML from SL classes at the milking stage, except the Green and RE-1 bands. In the ripening stage, RE-3, NIR-1, and NIR-2 could also distinguish ML from SL classes (Appendix Table 4).

Table 12: Kruskal Wallis test result for Sentinel-2 data

Lodging classes	P-value Blue	P-value Green	P-value Red	P-value RE-1	P-value RE-2	P-value RE-3	P-value NIR-1	P-value NIR-2	P-value SWIR-1	P-value SWIR-2
H-ML-SL	0.00	0.00	0.00	0.00	0.00	0.19	0.11	0.04	0.00	0.00

Table 13: Post hoc test result of Sentinel-2 data

Stem elongation to ripening										
Class Pairs	**P-value Blue**	**P-value Green**	**P-value Red**	**P-value RE-1**	**P-value RE-2**	**P-value RE-3**	**P-value NIR-1**	**P-value NIR-2**	**P-value SWIR-1**	**P-value SWIR-2**
H-ML	.37	.09	.37	.05	.31	Na	Na	.45	.04	.17
H-SL	.00	.00	.00	.00	.00	Na	Na	.01	.00	.00
ML-SL	.10	.02	.04	.02	.19	Na	Na	.40	.03	.03

3.5. Random Forest classification results

The Random Forest classifier, which was trained by using three features and five hundred decision tree of Sentinel-1 data, ten features and five hundred decision trees of Sentinel-2 data, and thirteen features and five hundred decision trees of Sentinel-1 and Sentinel-2 data combination, respectively, produces a confusion matrix and OOB error estimates (Table 14). For instance, the Random Forest classifier of the Sentinel-1 data (σVV°, σVH°, and σVH/VV°) misclassified the ML class into the other classes (He and SL); hence, the class error rate of ML class was higher. On the other hand, the overall error rate of (OOB) of the Sentinel-1 based Random Forest classifier was 16.27.

Table 14: Confusion matrix and OOB error rate of Random Forest classifier

	Sentinel-1				**Sentinel-2**				**Sentinel-1 and Sentinel-2**			
Classes	**He**	**ML**	**SL**	**Class error**	**He**	**ML**	**SL**	**Class error**	**He**	**ML**	**SL**	**Class error**
He	109	0	6	**0.05**	36	0	5	**0.12**	36	1	4	**0.12**
ML	5	3	3	**0.737**	5	0	3	**1.00**	2	2	4	**0.75**
SL	12	1	27	**0.33**	4	1	30	**0.14**	2	2	31	**0.11**
OOB estimate of error rate = 16.27%					**OOB estimate of error rate = 21.43%**				**OOB estimate of error rate = 17.86%**			

The Sentinel-2 data-based Random Forest classifier confusion matrix shows the ML class completely misclassified; thus, the class error rate of the ML class was 100%. However, since most of the He and SL classes were classified properly, their class error rate was moderate. The OOB error rate of the Sentinel-2 data-based Random Forest classifier was higher than the Sentinel-1 and combination of Sentinel-1 and Sentinel-2 data-based model because of the highest number of misclassified classes in the overall model.

The confusion matrix of the Sentinel-1 and Sentinel-2 combination data-based Random Forest classifier shows the class error rate of He and SL classes was lower than the ML class. On the other hand, the OOB error rate of the overall model was lower than the Sentinel-2 based Random Forest classifier due to fewer misclassified classes in the Sentinel-1 and Sentinel-2 combination-based Random Forest model.

Table 15 shows the most important features (variables) of Sentinel-1, Sentinel-2, and their combination data in the development of the Random Forest classifier. For example, σVH° and σVH/VV° were the first and the second most important variables in the development of Sentinel-1 data-based Random Forest classifier. In contrast, σVV° was the least important variable. The ranking was based on the mean decrease Gini impurity value for each variable. On the other hand, RE-1, SWIR-1, RE-2, Green, NIR-1 were the top five important Sentinel-2 variables in Sentinel-2 data-based Random Forest classifier development. Moreover, σVH°, SWIR-1, RE-1, Green, and RE-2 of Sentinel-1 and Sentinel-2 data, respectively, were the top five most important variables in the Random Forest classification of Sentinel-1 and Sentinel-2 combination data.

Table 15: Variable of importance for the Random Forest Classifier

Variable of importance	Sentinel-1			Sentinel-2			Sentinel-1 and Sentinel-2		
	Variable	**MeanDecreaseGini**	**Rank**	**Variable**	**MeanDecreaseGini**	**Rank**	**Variable**	**MeanDecreaseGini**	**Rank**
	VH	38.85	1	RE1	6.77	1	VH	5.38	1
	VH/VV	19.09	2	SWIR1	6.61	2	SWIR1	5.12	2
	VV	17.79	3	RE2	5.55	3	RE1	4.77	3
				Green	5.02	4	Green	4.73	4
				NIR1	4.40	5	RE2	4.09	5
				NIR2	4.24	6	Red	3.39	6
				SWIR2	3.90	7	VH/VV	3.17	7
				RE3	3.78	8	NIR2	3.17	8
				Red	3.69	9	NIR1	3.02	9
				Blue	3.37	10	RE3	2.99	10
							VV	2.93	11
							SWIR2	2.92	12
							Blue	2.28	13

The overall accuracy and kappa coefficient of the Random Forest model developed by the combination of Sentinel-1 and Sentinel-2 data was higher than Sentinel-1 and Sentinel-2 data alone. This implies the combination of optical and active sensor data could improve the predictive power of the Random Forest model. Moreover, Sentinel-2 has better predictive power than Sentinel-1 (Table 16).

Table 16: Random Forest classifier prediction outputs

	Sentinel-1					Sentinel-2					Sentinel-1 and Sentnel-2				
Class	**He**	**ML**	**SL**	**Total**	**UA**	**He**	**ML**	**SL**	**Total**	**UA**	**He**	**ML**	**SL**	**Total**	**UA**
He	38	1	3	42	**91**	15	1	0	16	**94**	16	1	0	17	**94**
ML	4	0	1	5	**0**	0	1	0	1	**100**	0	1	0	1	**100**
SL	3	1	11	15	**73**	3	3	13	19	**68**	2	3	13	18	**72**
Total	45	2	15	62		18	5	13			18	5	13		
PA (%)	**84**	**0**	**73**			**83**	**20**	**100**			**83**	**20**	**100**		
OA (%)	**79**					**80**					**83**				
Kappa	**0.53**					**0.67**					**0.71**				

3.5.1. Random Forest LS prediction map

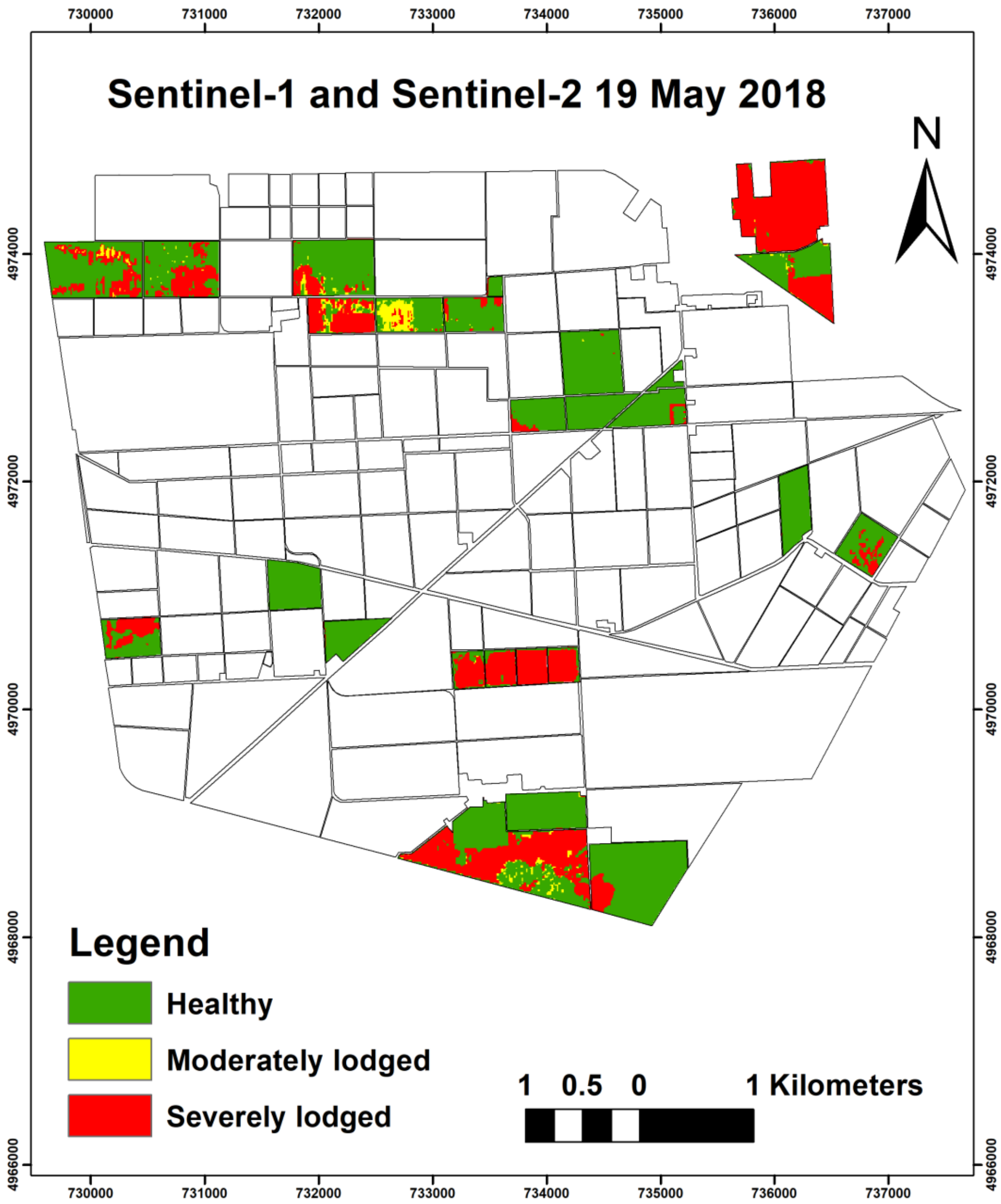


Figure 12: Bonifiche Ferraresi farm wheat LS map predicted from the combination of Sentinel-1 and Sentinel-2 data on 19 May 2018

Figure 12 shows the Random Forest LS classification map of 19 May 2018 Sentinel-1 and Sentinel-2 combination data. All the three wheat LS classes were appropriately represented.

4. DISCUSSION

In this study, we assessed the potential of the backscatter and spectral data of Sentinel-1, Sentinel-2, and their combination, respectively, in wheat lodging detection and classification using Random Forest classifier. Assessing the potentials of the three datasets is important in replacing the conventional lodging detection techniques, which are constrained mainly by their inherent bias, limited spatial coverage, and costs. Consequently, it may help farmers in the real-time monitoring of the phenomenon and to reduce yield losses. Besides, it will help to reduce extra harvesting and grain drying costs. Furthermore, it may help to solve disagreements between farmers and insurance companies regarding crop losses.

Firstly, we examined the performances of Sentinel-1 backscatter and Sentinel-2 spectral data in wheat lodging detection at different growth stages via backscatter and spectral analysis, respectively. Then we established the statistical relationship between the field-based LS and Sentinel-1 and Sentinel-2 data. Finally, we performed and compared Random Forest classification using Sentinel-1, Sentinel-2, and their combination, respectively. In the following sections, the most important findings of this research, limitations, and future works are discussed.

4.1. Backscatter and spectral behaviour of wheat

In the normal crop growth cycle, the backscattering properties of cereal crops are strongly affected by the crop phenology and associated plant and soil parameter changes (Larranaga et al., 2013). As mentioned in section 3.1, the σVV°, σVH°, and σVH/VV° of He wheat were higher at the stem elongation growth stage of wheat (Figure 6a). Wheat has a small plant canopy during its initial growth stages; therefore, the incoming wave interacts with the underlying soil surface and the σVV°, σVH°, and σVH/VV° are mostly driven by soil moisture and roughness of the soil (Chauhan, et al., 2020c; Dobson & Ulaby, 1981; Song & Wang, 2019). However, when wheat changed its growth stages to booting, flowering, and milking stages, the σVV°, σVH°, and σVH/VV° were mostly affected by LAI, fresh/dry biomass (FB/DB), and other physical parameters of the wheat plant. For example, at the booting stage and flowering stages of wheat, the σVV° and σVH° of wheat decreased (Figure 6b and Figure 7a). It could be due to the wave attenuation of wheat through their vertical structure, leaf area index, fresh/dry biomass (Chauhan, et al., 2020c; Mattia et al., 2003; Yang et al., 2015). In contrast, the increased σVH/VV° of wheat could be due to a canopy thickness (Song & Wang, 2019). Chauhan et al., (2020c) reported that the σVV°, σVH° of wheat decreased due to the increase of wheat biophysical and biochemical parameters. At the ripening stage, the leaf area index, fresh/dry biomass, and moisture content of wheat gradually decreased; hence the σVV°, σVH°, and σVH/VV° of wheat increased due to reduction of wave attenuation and increasing of soil backscatter.

When lodging occurred, the vertical structure of the wheat plant altered; hence the usual backscattering properties of wheat changed. As presented in Figure 7a and 7b, the mean σVV° and σVH°, and σVH/VV° of ML and SL wheat samples were higher than the He wheat samples at flowering and milking stages. The increment of the σVV° and σVH° of ML and SL wheat could be due to the loss of the vertical structure (the increasing of crop angle of inclination) of wheat subsequently the loss of wave attenuation and multiple scattering, respectively (Chauhan, et al., 2020a; Chauhan, et al., 2020c; Yang et al., 2015; Zhao et al., 2017). On the other hand, the σVH/VV° was sensitive to differentiate SL wheat from ML and He wheat at the milking growth stage it could be due to the strong sensitivity of σVH° and σVV° to crop structural changes at the milking stage (Chauhan, et al., 2020c). Yang et al., (2015) found that both σVV° and σHV° were sensitive to wheat lodging. In addition, Chauhan, et al., (2020c) reported σVH° and σVV° as the first and second most sensitive and important parameter for wheat lodging detection, respectively. The result of this study presented on section 3.1 were consistent with those of Yang et al., (2015), Chauhan, et al., (2020c), and Zhao et al., (2017).

Lodging stress affect crop biophysical and biochemical properties and cause changes in the spectral reflectance of vegetation canopies (Chauhan et al., 2019b; Chauhan et al., 2020c). As shown in section 3.2, the spectral reflectance of ML and SL wheat samples was higher than He wheat samples in all spectral regions (Figure 9b, 10a, and 10b). The change in the crop structure (an increase of CAI and leaf surface area) and biochemical parameters (the reduction of chlorophyll and water concentration) could lead to the raising of ML and SL spectral reflectance in the visible (blue, green, red), RE (RE-1, RE-2, RE-3), NIR (NIR-1, NIR-2), and SWIR (SWIR-1, SWIR-2) regions of the spectrum (Chauhan et al., 2020c; Chauhan et al., 2019b; Setter et al., 1997; Wang et al., 2020). Chauhan et al., (2019b), Chauhan, et al., (2020c), Wang et al., (2020) also found that lodging increase the magnitude of wheat spectral reflectance in all spectral regions of Sentinel-2.

The Kruskal Wallis and the post hoc test results of Sentinel-1 (σVV°) and Sentinel-2 data (Green, Red, RE-1, SWIR-1, SWIR-2, Blue, RE-2, and NIR-2) showed a significant difference in terms of differentiating He and lodged groups (ML and SL). However, some Sentinel-1 (σVH° and σVH/VV°) and Sentinel-2 (RE-3 and NIR-1) parameters showed a non-significant test value which was contradicting with Figure 7a, 7b, 8, 9b, 10a, and 10b. We pooled the SL and very severely lodged (VSL) wheat LS groups of previous Chauhan, et al., (2020c) work (Appendix Table 5); hence the pooling of the two LS groups could make the σVH°, σVH/VV°, RE-3, NIR-1 bands non-significant (Table 12). In addition, the maximum NIR-1 spectral reflectance of He and SL wheat samples was almost similar at flowering and ripening stages of wheat (Figure 9b and 10b); thus, it could create a problem for the Kruskal Wallis and post hoc tests to properly differentiate the three wheat LS groups. Furthermore, the Kruskal Wallis and post hoc tests use mean rank-sum to identify the significant difference between the three LS groups (He, ML, and SL), whereas visual interpretation was used to interpret the box plots of Sentinel-1 data, so the difference between the two criteria may create confusion in the interpretation.

4.2. Random Forest classification

The OOB error rate of the Random Forest classifier, which was trained by using Sentinel-1 data (σVV°, σVH°, and σVH/VV°) was moderate due to high error of one of the three classes. It could be due to the confusing plant orientation of the ML class. Since ML samples have slightly inclined stems, it could attenuate some of the VV polarized incoming waves through their inclined stems (Yang et al., 2015). Therefore, it could create confusion for the model to distinguish the ML class from the He and SL classes (Figure 7a and 7b) and contribute higher for the overall OOB error rate of the model. σVH° was the most important Sentinel-1 variable in the LS classification using Random Forest classifier. Lodging changes the vertical structure of wheat; hence, it would result in multiple scattering of wheat. These backscatter changes could easily be detected through σVH° because of its inherent sensitivity for multiple backscattering of wheat during lodging (Chauhan, et al., 2020a; Chauhan, et al., 2020c; Yang et al., 2015; Wang et al., 2020). Wang et al., (2020) also found that σVH° was the most important variable in the Random Forest LS classification of Sentinel-1 data. However, the result of the Random Forest classifier variable of importance function and the Kruskal Wallis and post hoc test were not consistent. It could be due to the difference in the evaluation criteria between the Random Forest classifier variable of importance function and Kruskal Wallis and post hoc tests. The Random Forest classifier variable of importance function uses the Mean Decrease Gini

(Mean Decrease Impurity) index as a criterion for ranking the variables in the classification. Mean Decrease Gini measures the importance of features in decreasing impurities in the Random Forest decision tree training. On the other hand, Kruskal Wallis and post hoc tests use mean rank-sum to identify the potential of Sentinel-1 variables in the differentiation of He, ML, and SL wheat LS groups.

The Sentinel-2 data-based Random Forest model OOB error rate was higher than the Sentinel-1 and the combination data-based Random Forest models (Table 14). It could be due to the misclassification of the ML classes into SL and He classes because of the close average spectral reflectance value of ML classes with He class at flowering stage, and ML class with SL class at milking stage, respectively (Figure 9b and 10a). However, the class error rate of He and SL classes was low due to a clear difference in He and SL classes' spectral reflectance in most spectral bands (Figure 9b, Figure 10a, and 10b). Although the OOB error rate of the Sentinel-2 data-based Random Forest model was higher, the predictive power of the Sentinel-2 data-based Random Forest model was still slightly higher than the Sentinel-1 data-based Random Forest model. It could be due to the difference in spectral and spatial resolution of Sentinel-1 and Sentinel-2 datasets. The top five most important variables of Sentinel-2 in the Random Forest model-based wheat LS classification were RE-1, SWIR1, RE-2, Green, and NIR1. The lodging stress causes a reduction of chlorophyll and water concentration of wheat; hence, it would increase reflectance in the visible and RE regions. Therefore, these spectral changes could easily be captured through Green, RE-1, and RE-2 bands of Sentinel-2 since they have inherent sensitivity to reflectance changes due to chlorophyll and water changes (Chauhan et al., 2019b; Chauhan et al., 2020c; Wang et al., 2020). Furthermore, the structural changes caused by lodging could also increase reflectance in the NIR region of the spectrum; thus, the NIR-1 band of Sentinel-2 could easily capture the spectral changes. The results were similar to Chauhan et al., (2019b), Chauhan, et al., (2020c), and Wang et al., (2020) findings, even though there were slight differences.

The Kruskal Wallis and post hoc test results also confirmed that the RE-1, SWIR-1, RE-2, and Green bands could distinguish He and SL classes (Table 12 and 13). But some of the Random Forest variables of importance outputs were not consistent with the Kruskal Wallis and post hoc test results. The difference between the parameter used in the Random Forest classifier and Kruskal Wallis and post hoc tests could be a reason for the difference between the rank of the variables of importance for the Random Forest-based Sentinel-2 data classification.

In the combination of Sentinel-1 and Sentinel-2 data-based Random Forest model, the class error rate of ML class and the OOB error rate of the model was slightly improved compared to the Sentinel-2 based Random Forest model (Table 14). This indicates the combination of SAR and optical features could play a role in the reduction of the class-specific and overall error rate of the model (Wang et al., 2020). Wang et al., (2020) also found an improved OOB error rate using the combination of Sentinel-1 and Sentinel-2 data. The $\sigma VH°$ of Sentinel-1 data was the most important variable in the combination data-based Random Forest model because of its sensitivity to lodging-induced structural changes. The other four important variables in the combination data-based Random Forest model were SWIR-1, RE-1, Green, and RE-2. These Sentinel-2 parameters were sensitive to crop structural and biochemical changes. Therefore, the combination of the Sentinel-1 data, which can acquire without any cloud effect with cloud-sensitive Sentinel-2 data, provides better accuracy.

4.3. Did the combination of Sentinel-1 and Sentinel-2 data improve classification accuracy?

The LS prediction output of the combination data-based Random Forest model was higher (83%) than the individual Sentinel-1 (79%) and Sentinel-2 (80%) data-based Random Forest model prediction outputs. The

Sentinel-1 backscatter features are highly sensitive to the structural changes of crops. On the other hand, Sentinel-2 features are sensitive to the crop biochemical and biophysical changes. Therefore, combining the features from the two datasets would led to a better LS classification of wheat. Wang et al., (2020) also reported a maximum prediction accuracy (91.29%) in rice crop lodging classification using a Random Forest model with Sentinel-1 and Sentinel-2 combination data.

4.4. Limitations and future works

The limitation of this study and areas for future improvement are presented as follows. In this study, there was a big sample size difference between He, ML, and SL wheat; unbalanced data could lead the Random Forest model to be biased to the class with maximum sample size (Bader-El-Den et al., 2019; Galar et al., 2012). In addition, it also affects the results of Kruskal Wallis and post hoc tests. Therefore, future work should use balanced field and remote sensing data. The other major limitation of the study was the gap between the field and remote sensing data acquisition, especially for Sentinel-2 data which was greater than six days. Therefore, a smaller time interval between field data collection and remote sensing data acquisition would help when relating plants' biophysical and biochemical changes to remote sensing observations. Appropriately acquired data would also help to train the Random Forest model properly.

The combination of the two spaceborne datasets (Sentinel-1 and Sentinel-2) provided an improved LS classification accuracy. However, further investigations on the integration of full polarimetric, high-resolution short (X) or long (L) band SAR data with freely available Sentinel-2 data could provide better lodging detection and classification accuracy. The full polarimetric SAR data provides polarimetric decomposition information. In addition, the L-band SAR data can penetrate the plant canopy and acquire the data from the ground, whereas X-band SAR datasets acquire information from the top of the crop canopy, so the investigations of these two different bands SAR data by combining with optical data could provide another information's in terms of lodging detections.

The Random Forest model which was used to classify LS based on Sentinel-1, Sentinel-2, and their combinations datasets, performed very well. However, further investigations should be performed to check the performances of different machine learning algorithms, such as deep learning algorithms in the LS classification of the three datasets.

The study demonstrated the performance of Sentinel-1, Sentinel-2, and the combination of Sentinel-1 and Sentinel-2 data in wheat lodging detection and classification in a specific area. It will be a good addition if future research focuses on investigating the potential of the three datasets in large areas with different agro-ecologies.

4.5. Implications of the study

The main purpose of this research was to evaluate the potential of Sentinel-1, Sentinel-2, and their combination data in wheat lodging detection and classifications using Random Forest classifier. Our result demonstrated that the integration of the two datasets could improve the classification accuracy of the Random Forest model. Therefore, this study could be considered as a promising prospect in the gradual replacement of the conventional lodging monitoring techniques with effective and efficient remote sensing techniques. Remote sensing techniques in lodging monitoring provide accurate and precise information regarding the phenomenon; hence, it will help farmers and other respective stakeholders make appropriate decisions. For instance, farmers can monitor the health of their crops remotely, it will help them to maintain the quality and quantity of their crop yield higher, and it would help them to reduce the cost incurred for harvesting, drying, and labour. Furthermore, remote sensing-based lodging monitoring would help insurance companies to get evidence about the phenomenon and give appropriate remedial actions. On the other hand, the Random Forest model used in this study would help the remote sensing communities to correctly classify the remote sensing data with a short period of time and energy; hence, it reduces the power and energy required for making manual classifications.

5. CONCLUSION

The study examined the performances of Sentinel-1, Sentinel-2, and their combination data in wheat lodging detection and classification with a Random Forest classifier. The following conclusions were drawn based on statistical analysis and a Random Forest classification using the three datasets. The statistical summary of Sentinel-1 backscatter data revealed that Sentinel-1 could effectively distinguish the lodged wheat classes (ML and SL) from the healthy (He) wheat class, and the Random Forest LS classification result showed an overall accuracy and kappa coefficient of 79% and 0.53, respectively. This indicates the usefulness of Sentinel-1 data in the operational wheat lodging detection and classification. Moreover, the Random Forest LS classification result of Sentinel-1 data showed that $\sigma VH°$ was the most important parameter for classifying healthy and lodged classes.

The investigation of Sentinel-2 spectral data showed an encouraging result. Most of the Sentinel-2 spectral bands could differentiate the lodged classes from the healthy class. Furthermore, the overall classification accuracy and kappa coefficient of the Random Forest model using Sentinel-2 spectral bands was 80% and 0.65, respectively. Moreover, from the Sentinel-2 spectral bands, RE-1 was the most important variable in terms of differentiating the three LS classes. The other top five most important variables of Sentinel-2 in the Random Forest LS classification were SWIR-1, RE-2, Green, and NIR-1.

The overall accuracy and kappa coefficient of the combination of Sentinel-1 and Sentinel-2 data-based Random Forest model was 83% and 0.71, respectively, which is higher than the individual Sentinel-1 and Sentinel-2 overall accuracy and kappa coefficient, respectively. This implies the combination of the two datasets provides a good opportunity to improve the overall performance of the combination-based Random Forest model. Furthermore, the $\sigma VH°$, SWIR-1, RE-1 bands of Sentinel-1 and Sentinel-2 were the top three important variables in the combination data random forest classification.

REFERENCES

AHDB. (2005). Lodging in winter wheat. https://doi.org/https://projectblue.blob.core.windows.net/media/Default/Imported%20Publication%20Docs/Avoiding%20lodging%20in%20winter%20wheat%20-%20practical%20guidelines.pdf

Atzberger, C. (2013). Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sensing*, *5*(2), 949–981. https://doi.org/10.3390/rs5020949

Bader-El-Den, M., Teitei, E., & Perry, T. (2019). Biased Random Forest for Dealing with the Class Imbalance Problem. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(7), 2163–2172. https://doi.org/10.1109/TNNLS.2018.2878400

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Berry, P. M., & Spink, J. (2012). Predicting yield losses caused by lodging in wheat. *Field Crops Research*, *137*, 19–26. https://doi.org/10.1016/j.fcr.2012.07.019

Berry, P. M., Sterling, M., Baker, C. J., Spink, J., & Sparkes, D. L. (2003). A calibrated model of wheat lodging compared with field measurements. *Agricultural and Forest Meteorology*, *119*(3–4), 167–180. https://doi.org/10.1016/S0168-1923(03)00139-4

Bouman, B.A.M. (1991). Crop parameter estimation from ground-based x-band (3-cm wave) radar backscattering data. *Remote Sensing of Environment*, *37*(3), 193–205. https://doi.org/10.1016/0034-4257(91)90081-G

Bouman, B.A.M, & van Kasteren, H. W. J. (1990). Ground-based X-band (3-cm wave) radar backscattering of agricultural crops. II. Wheat, barley, and oats; the impact of canopy structure. *Remote Sensing of Environment*, *34*(2), 107–119.

Breiman, L. (2001). ST4_Method_Random_Forest. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1017/CBO9781107415324.004

Chauhan, S., Darvishzadeh, R., Lu, Y., Stroppiana, D., Boschetti, M., Pepe, M., & Nelson, A. (2019b). Wheat lodging assessment using multispectral uav data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *42*(2/W13), 235–240. https://doi.org/10.5194/isprs-archives-XLII-2-W13-235-2019

Chauhan, S., Darvishzadeh, R., Boschetti, M., & Nelson, A. (2020b). Discriminant analysis for lodging severity classification in wheat using RADARSAT-2 and Sentinel-1 data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *164*(March), 138–151. https://doi.org/10.1016/j.isprsjprs.2020.04.012

Chauhan, S., Darvishzadeh, R., Boschetti, M., & Nelson, A. (2020a). Estimation of crop angle of inclination for lodged wheat using multi-sensor SAR data. *Remote Sensing of Environment*, *236*(November 2019), 111488. https://doi.org/10.1016/j.rse.2019.111488

Chauhan, S., Darvishzadeh, R., Boschetti, M., Pepe, M., & Nelson, A. (2019a). Remote sensing-based crop lodging assessment: Current status and perspectives. *ISPRS Journal of Photogrammetry and Remote Sensing*, *151*(March), 124–140. https://doi.org/10.1016/j.isprsjprs.2019.03.005

Chauhan, S., Darvishzadeh, R., Lu, Y., Boschetti, M., & Nelson, A. (2020c). Understanding wheat lodging using multi-temporal Sentinel-1 and Sentinel-2 data. *Remote Sensing of Environment*, *243*, 111804. https://doi.org/10.1016/j.rse.2020.111804

Colomina, I., Blázquez, M., Molina, P., Parés, M. E., & Wis, M. (2008). Towards a new paradigm for high-resolution low-cost photogrammetryand remote sensing. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *37*. https://www.isprs.org/proceedings/xxxvii/congress/1_pdf/205.pdf

Constantinescu, C. A., Herbei, M. V, Manea, D., & Sala, F. (2017). *Analysis of Some Deficiencies in Crops of Wheat and Barley Based on Terrestrial and Aerial Images*. *49*(1998), 95–103. https://scholar.google.nl/scholar?hl=en&as_sdt=0%252C5&q=Constantinescu%252C+C.+A.%252C+Herbei%252C+M.+V%252C+Manea%252C+D.%252C+%2526+Sala%252C+F.+%25282017

%2529.+Analysis+of+Some+Deficiencies+in+Crops+of+Wheat+and+Barley+Based+on+Terrestrial+and+Aerial+Images.+49%25281998%2529%252
Dobson, M. C., & Ulaby, F. (1981). Microwave Backscatter Dependence on Surface Roughness, Soil Moisture, and Soil Texture: Part III—Soil Tension. *IEEE Transactions on Geoscience and Remote Sensing*, *GE-19*(1), 51–61. https://doi.org/10.1109/TGRS.1981.350328
Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, *120*, 25–36. https://doi.org/10.1016/j.rse.2011.11.026
Fang, Z., & Cao, C. (2014). Estimation of Forest Canopy Height Over Mountainous Areas Using Satellite Lidar. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(7), 3157–3166. https://doi.org/10.1109/JSTARS.2014.2300145
Fernández-Manso, A., Fernández-Manso, O., & Quintano, C. (2016). SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity. *International Journal of Applied Earth Observation and Geoinformation*, *50*, 170–175. https://doi.org/10.1016/j.jag.2016.03.005
Fitch, B. W., Walraven, R. L., & Bradley, D. E. (1984). Polarization of light reflected from grain crops during the heading growth stage. *Remote Sensing of Environment*, *15*(3), 263–268. https://doi.org/10.1016/0034-4257(84)90036-1
Galar, M., Fern, A., Barrenechea, E., & Bustince, H. (2012). *Hybrid-Based Approaches*. *42*(4), 463–484.
Gao, B.-C. (1996). Naval Research Laboratory, 4555 Overlook Ave. *Remote Sens. Environ*, *7212*(April), 257–266. https://doi.org/https://cpb-us-w2.wpmucdn.com/sites.udel.edu/dist/d/1835/files/2014/06/NDWI-A-Normalized-Difference-Water-Index-for-Remote-Sensing-of-Vegetation-Liquid-Water-From-Space-1ko95nn.pdf
Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, *27*(4), 294–300. https://doi.org/10.1016/j.patrec.2005.08.011
Hijmans, R. J. (2020). *raster: Geographic Data Analysis and Modeling. R package version 3.4-5*. https://cran.r-project.org/package=raster
Horning, N. (2010). Random Forests: An algorithm for image classification and generation of continuous fields data sets. *International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences 2010*, 1–6. https://doi.org/http://wgrass.media.osaka-cu.ac.jp/gisideas10/papers/04aa1f4a8beb619e7fe711c29b7b.pdf
Huete, A. R. (1988). A Soil-Adjusted Vegetation Index (SAVI). *Remote Sensing Of Environment 25(3):295-309*.
Huete, A. R., Didan, K., & Van Leeuwen, W. (1999). Modis Vegetation Index. *Vegetation Index and Phenology Lab*, *3(1)*(April 1999), 129.
Kuhn, M. (2020). *caret: Classification and Regression Training*. https://cran.r-project.org/package=caret
Larranaga, A., Alvarez-Mozos, J., Albizua, L., & Peters, J. (2013). Backscattering behavior of rain-fed crops along the growing season. *IEEE Geoscience and Remote Sensing Letters*, *10*(2), 386–390. https://doi.org/10.1109/LGRS.2012.2205660
Li, Z., Chen, Z., Wang, L., Liu, J., & Zhou, Q. (2014). Area extraction of maize lodging based on remote sensing by small unmanned aerial vehicle. *Transactions of the Chinese Society of Agricultural Engineering*, *30*(19), 207–213.
Liu, H. Y., Yang, G. J., & Zhu, H. C. (2014). The extraction of wheat lodging area in UAV's image used spectral and texture features. In *Applied Mechanics and Materials* (Vols. 651–653). https://doi.org/10.4028/www.scientific.net/AMM.651-653.2390
MacFarland, T. W., & Yates, J. M. (2016). Introduction to Nonparametric Statistics for the Biological Sciences Using R. In *Introduction to Nonparametric Statistics for the Biological Sciences Using R*. https://doi.org/10.1007/978-3-319-30634-6
Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M. E., García-Santos, G., Fernandes, R., & Berger, M. (2012). Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sensing of Environment*, *120*, 91–101. https://doi.org/10.1016/j.rse.2011.09.026
Mattia, F., Le Toan, T., Picard, G., Posa, F. I., D'Alessio, A., Notarnicola, C., Gatti, A. M., Rinaldi, M., Satalino, G., & Pasquariello, G. (2003). Multitemporal C-band radar measurements on wheat fields. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(7 PART I), 1551–1560. https://doi.org/10.1109/TGRS.2003.813531
Moran, M. S., Inoue, Y., & Barnes, E. M. (1997). Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing of Environment*, *61*(3), 319–346.

https://doi.org/10.1016/S0034-4257(97)00045-X
Niu, L., Feng, S., Ding, W., & Li, G. (2016). Influence of speed and rainfall on large-scale wheat lodging from 2007 to 2014 in China. *PLoS ONE*, *11*(7), 1–15. https://doi.org/10.1371/journal.pone.0157677
Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217–222. https://doi.org/10.1080/01431160412331269698
Pearson, R. L., & Miller, L. D. (1972). Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie. *Rse*, 1355. https://doi.org/https://ui.adsabs.harvard.edu/abs/1972rse..conf.1355P/abstract
Pebesma, E., & Bivand, R. S. (2005). S classes and methods for spatial data: the sp package. *R News*, *5*(2), 9–13.
Phillips, R. C., & Ward, J. K. (2020). Comparison of Machine Learning Algorithms to Detect Crop Lodging using UAS Imagery. In *2020 ASABE Annual International Virtual Meeting* (p. 1). ASABE. https://doi.org/https://doi.org/10.13031/aim.202001034
Pinthus, M. J. (1974). *Lodging in Wheat, Barley, and Oats: The Phenomenon, its Causes, and Preventive Measures* (N. C. B. T.-A. in A. Brady (ed.); Vol. 25, pp. 209–263). Academic Press. https://doi.org/https://doi.org/10.1016/S0065-2113(08)60782-8
Rawson, H. ., & Macpherson, H. G. (2000). Irrigated wheat. In *Food and Agriculture Organization of the United Nations*. http://www.fao.org/3/x8234e/x8234e00.htm#Contents
Richardson, A. J., & Wiegand, C. L. (1977). Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing*, *43*(12), 1541–1552.
Rouse, J. W. (1974). Monitoring the vernal advancement of retrogradation of natural vegetation, NASA/GSFG, Type III. *Final Report*, *371*. https://doi.org/https://ci.nii.ac.jp/naid/10005114879/
Rowlingson, R. B. and T. K. and B. (2021). *rgdal: Bindings for the "Geospatial" Data Abstraction Library*. https://cran.r-project.org/package=rgdal
Setter, T. L., Laureles, E. V., & Mazaredo, A. M. (1997). Lodging reduces yield of rice by self-shading and reductions in canopy photosynthesis. *Field Crops Research*, *49*(2–3), 95–106. https://doi.org/10.1016/S0378-4290(96)01058-1
Shu, M., Zhou, L., Gu, X., Ma, Y., & Sun, Q. (2020). ScienceDirect Monitoring of maize lodging using multi-temporal Sentinel-1 SAR data. *Advances in Space Research*, *65*(1), 470–480. https://doi.org/10.1016/j.asr.2019.09.034
Shu, M., Zhou, L., Gu, X., Ma, Y., Sun, Q., Yang, G., & Zhou, C. (2020). Monitoring of maize lodging using multi-temporal Sentinel-1 SAR data. *Advances in Space Research*, *65*(1), 470–480. https://doi.org/10.1016/j.asr.2019.09.034
Song, Y., & Wang, J. (2019). Mapping winter wheat planting area and monitoring its phenology using Sentinel-1 backscatter time series. *Remote Sensing*, *11*(4). https://doi.org/10.3390/rs11040449
Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B. Ö., Floury, N., Brown, M., Traver, I. N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., … Rostan, F. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment*, *120*, 9–24. https://doi.org/10.1016/j.rse.2011.05.028
Vargas, J. Q., Khot, L. R., Peters, R. T., Chandel, A. K., Molaei, B., & Details, A. F. S. (2020). *Low Orbiting Satellite and Small UAS-Based High-Resolution Imagery Data to Quantify Crop Lodging : A Case Study in Irrigated Spearmint*. *17*(5), 755–759. https://doi.org/https://ieeexplore.ieee.org/abstract/document/8822470
Vescovo, L., Gianelle, D., Dalponte, M., Miglietta, F., Carotenuto, F., & Torresan, C. (2016). Hail defoliation assessment in corn (Zea mays L.) using airborne LiDAR. *Field Crops Research*, *196*, 426–437. https://doi.org/http://dx.doi.org/10.1016/j.fcr.2016.07.024
Wang, J., Li, K., Shao, Y., Zhang, F., Wang, Z., Guo, X., Qin, Y., & Liu, X. (2020). Analysis of combining SAR and optical optimal parameters to classify typhoon-invasion lodged rice: a case study using the random forest method. *Sensors (Switzerland)*, *20*(24), 1–18. https://doi.org/10.3390/s20247346
Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org
Wiener, A. L. and M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22. https://cran.r-project.org/doc/Rnews/
Xiang, D. B., Zhao, G., Wan, Y., Tan, M. L., Song, C., & Song, Y. (2016). Effect of planting density on lodging-related morphology, lodging rate, and yield of tartary buckwheat (Fagopyrum tataricum). *Plant Production Science*, *19*(4), 479–488. https://doi.org/10.1080/1343943X.2016.1188320

Yang, H., Chen, E., Li, Z., Zhao, C., Yang, G., Pignatti, S., Casa, R., & Zhao, L. (2015). Wheat lodging monitoring using polarimetric index from RADARSAT-2 data. *International Journal of Applied Earth Observation and Geoinformation*, *34*(1), 157–166. https://doi.org/10.1016/j.jag.2014.08.010

Zhang, Z., Flores, P., Igathinathane, C., Naik, D. L., Kiran, R., & Ransom, J. K. (2020). *Wheat Lodging Detection from UAS Imagery Using Machine Learning Algorithms*. https://doi.org/https://www.mdpi.com/2072-4292/12/11/1838

Zhao, L., Yang, J., Li, P., Shi, L., & Zhang, L. (2017). Characterizing lodging damage in wheat and canola using radarsat-2 polarimetric SAR data. *Remote Sensing Letters*, *8*(7), 667–675. https://doi.org/10.1080/2150704X.2017.1312028

Zhou, L., Cheng, S., Sun, Q., Gu, X., Yang, G., Shu, M., & Feng, H. (2020). Remote sensing of regional-scale maize lodging using multitemporal GF-1 images. *Journal of Applied Remote Sensing*, *14*(01), 1. https://doi.org/10.1117/1.jrs.14.014514

APPENDICES

Table 1: Kruskal Wallis test results of Sentinel-1 data

Growth stage	Sentinel-1		
	P-value $\sigma VH°$	P-value $\sigma VV°$	P-value $\sigma VH/VV°$
Flowering	.07	.10	.09
Milking	.00	.00	.00
Ripening	.62	.14	.62

Table 2: Post hoc test pairwise comparison results of Sentinel-1 data

Class Pairs	Flowering			Milking			Ripening		
	P-value $\sigma VH°$	P-value $\sigma VV°$	P-value $\sigma VH/VV°$	P-value $\sigma VH°$	P-value $\sigma VV°$	P-value $\sigma VH/VV°$	P-value $\sigma VH°$	P-value $\sigma VV°$	P-value $\sigma VH/VV°$
He-ML	Na	Na	Na	.04	.01	.40	Na	Na	Na
He-SL	Na	Na	Na	.00	.00	.00	Na	Na	Na
ML-SL	Na	Na	Na	.13	.40	.06	Na	Na	Na

Table 3: Kruskal Wallis test results of Sentinel-2 data

Growth stage	Sentinel-2 Bands									
	P-value Blue	P-value Green	P-value Red	P-value RE1	P-value RE2	P-value RE3	P-value NIR1	P-value NIR2	P-value SWIR1	P-value SWIR2
Flowering	0.11	0.10	0.12	0.12	0.10	0.12	0.08	0.05	0.12	0.14
Milking	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ripening	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 4: Post hoc test pairwise comparison results of Sentinel-2 data

Growth-stage	Class-Pairs	Sentinel-2 Bands									
		P-value Blue	P-value Green	P-value Red	P-value RE1	P-value RE2	P-value RE3	P-value NIR1	P-value NIR2	P-value SWIR1	P-value SWIR2
Flowering	He-ML	Na	Na	Na	Na	Na	Na	Na	.17	Na	Na
	He-SL	Na	Na	Na	Na	Na	Na	Na	.02	Na	Na
	ML-SL	Na	Na	Na	Na	Na	Na	Na	.45	Na	Na
Milking	He-ML	.11	.06	.18	.06	.21	.54	.41	.55	.26	.32
	He-SL	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	ML-SL	.04	.06	.03	.06	.02	.00	.01	.00	.02	.01
Ripening	He-ML	.12	.14	.10	.13	.25	.28	.37	.34	.19	.14
	He-SL	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	ML-SL	.18	.13	.19	.16	.06	.05	.03	.04	.06	.09

Table 5: Kruskal Wallis test results of Sentinel-2 data with four LS groups (stem elongation to ripening)

Test Statistics										
	Blue	Green	Red	RE1	RE2	RE3	NIR1	NIR2	SWIR1	SWIR2
Kruskal-Wallis H	16.603	41.138	22.667	48.633	19.980	7.800	8.944	11.192	46.340	31.128
df	3	3	3	3	3	3	3	3	3	3
Asymp. Sig.	.001	.000	.000	.000	.000	.050	.030	.011	.000	.000

```matlab
%%% MATLAB codes used to plot the average spectral reflectance of
wheat at different growth stages%%%

%Average Spectral reflectance of healthy wheat at stem elongation
and booting stages
StemH_average=mean(StemHealth);
BootingH_average=mean(BootHealth);

% To draw average spectral reflectance
hold on
aa=plot(wl,StemH_average, 'g.-');
bb=plot(wl,BootingH_average, 'm.-');
legend([aa, bb], 'Stem elongation He', 'Booting He')
hold off

% Flowering average spectral reflectance of wheat
FloweringH_average=mean(FlowerHealth);
FloweringML_average=mean(FlowerML);
FloweringSL_average=mean(FlowerSL);

% To draw average spectral reflectance
hold on
aa=plot(wl,FloweringH_average, 'g.-');
bb=plot(wl,FloweringML_average, 'k.-');
cc=plot(wl,FloweringSL_average, 'r.-');
legend([aa, bb, cc], 'He', 'ML', 'SL');
hold off

% Milking average spectral reflectance of wheat
MilkingH_average=mean(MilkHealth);
MilkinggML_average=mean(MilkML);
MilkingSL_average=mean(MilkSL);

% To plot the average spectral reflectance of wheat
hold on
aa=plot(wl,MilkingH_average, 'g.-');
bb=plot(wl,MilkinggML_average, 'k.-');
cc=plot(wl,MilkingSL_average, 'r.-');
legend([aa, bb, cc], 'He', 'ML', 'SL');
hold off

% Ripening average spectral reflectance of wheat
RipeningH_average=mean(RipeHealth);
RipeningML_average=mean(RipeML);
RipeningSL_average=mean(RipeningSL);

% To plot the average spectral reflectance of wheat
hold on
aa=plot(wl,RipeningH_average, 'g.-');
bb=plot(wl,RipeningML_average, 'k.-');
cc=plot(wl,RipeningSL_average, 'r.-');
legend([aa, bb, cc], 'He', 'ML', 'SL');
hold off
```

```r
### Codes used for the Random Forest classification of Sentinel-1 an
d Sentinel-2 data combination###

# only install if needed
if (!require("raster")) install.packages("raster")

if (!require("sf")) install.packages("sf")

if (!require("rgdal")) install.packages("rgdal")

if (!require("randomForest")) install.packages("randomForest")

if (!require("caret")) install.packages("caret")

# call libraries/packages
library(raster)      # to add raster pachage
library(sf)
library(rgdal)
library(randomForest)
library(caret)

getwd()

## To set working directory##
setwd("D:/My MSc. Proposal and thesis Documents/RS&FD/Sentinel1&Sent
inel2/S1&2_to_be-submmited/CSV")

## To add the CSV file of Sentinel-1 and Sentinel-2 data ombination#
#
LodgingscoreS1and2first_SLL<-read.csv('RefandBackscatter_SLL.csv')

## To remove the unnecessary columns from the data##
d<-within(LodgingscoreS1and2first_SLL, rm("X","Y"))

## To view the Sentinel-1 and Sentinel-2 combination data##
View(d)

# To split the data into He, ML and SL groups##
LS1_S1and2 <- d[which(d$Lodging.score==1), ]
LS2_S1and2 <- d[which(d$Lodging.score==2), ]
LS3_S1and2 <- d[which(d$Lodging.score==3), ]



## To split the He, ML, SL groups into 70/30 percent##
sample

set.seed(123)
LS1id_S1and2<-sample(2, nrow(LS1_S1and2), prob = c(0.7,0.3),replace
= TRUE)
LodgingscoreS1and2_trainfirst_SLL_LS1<-LS1_S1and2[LS1id_S1and2==1,]
LodgingscoreS1and2_testfirst_SLL_LS1<-LS1_S1and2[LS1id_S1and2==2,]

LS2id_S1and2<-sample(2, nrow(LS2_S1and2), prob = c(0.7,0.3),replace
= TRUE)
LodgingscoreS1and2_trainfirst_SLL_LS2<-LS2_S1and2[LS2id_S1and2==1,]
```

```r
LodgingscoreS1and2_testfirst_SLL_LS2<-LS2_S1and2[LS2id_S1and2==2,]

LS3id_S1and2<-sample(2, nrow(LS3_S1and2), prob = c(0.7,0.3),replace
= TRUE)
LodgingscoreS1and2_trainfirst_SLL_LS3<-LS3_S1and2[LS3id_S1and2==1,]
LodgingscoreS1and2_testfirst_SLL_LS3<-LS3_S1and2[LS3id_S1and2==2,]

## To merge the training and test samples of He, ML, and SL classes
and to get the data needed for random forest##
LodgingscoreS1and2_trainfirst_SLVSL <- rbind (LodgingscoreS1and2_tra
infirst_SLL_LS1,LodgingscoreS1and2_trainfirst_SLL_LS2,LodgingscoreS1
and2_trainfirst_SLL_LS3)
LodgingscoreS1and2_testfirst_SLVSL  <- rbind (LodgingscoreS1and2_tes
tfirst_SLL_LS1,LodgingscoreS1and2_testfirst_SLL_LS2,LodgingscoreS1an
d2_testfirst_SLL_LS3)

## To save the Lodging score (LS) as a factor variable##
LodgingscoreS1and2_trainfirst_SLVSL$Lodging.score<-as.factor(Lodging
scoreS1and2_trainfirst_SLVSL$Lodging.score)
LodgingscoreS1and2_testfirst_SLVSL$Lodging.score<-as.factor(Lodgings
coreS1and2_testfirst_SLVSL$Lodging.score)

## To select the optimal mtry from the training samples##
bestmtrys1<-tuneRF(LodgingscoreS1and2_trainfirst_SLVSL,LodgingscoreS
1and2_trainfirst_SLVSL$Lodging.score,stepFactor = 1.2,improve = 0.0
1,trace = T,plot = T)

## To run the Sentinel-1 and Sentinel-2 data combination Random Fore
st classification##
LodgingscoreRFS1and2_SLL<-randomForest(Lodging.score~., data = Lodgi
ngscoreS1and2_trainfirst_SLVSL, mtry=3, ntree=500)     #The accuracy
will

## To view the Random Forest classification outputs##
LodgingscoreRFS1and2_SLL

## To see the most important variables for the Random Forest classif
ication of Sentinel-1 and Sentinel-2 data combinations##
importance(LodgingscoreRFS1and2_SLL)

## To plot the most important variables in descending order##
varImpPlot(LodgingscoreRFS1and2_SLL)

## To run the Random Forest prediction##
predict_LodgingscoreS1and2_SLL<-predict(LodgingscoreRFS1and2_SLL, ne
wdata = LodgingscoreS1and2_testfirst_SLVSL, type = "class")

## To view the Random Forest prediction outputs##
predict_LodgingscoreS1and2_SLL

## To compare the prediction results by using of caret library##
confusionMatrix(table(predict_LodgingscoreS1and2_SLL,LodgingscoreS1a
nd2_testfirst_SLVSL$Lodging.score))
```

```
### To predict with Sentinel-1 and Sentinel-2 image combination ##


## To Read multi-band image data##
S1and2_19may_image_SLL<-stack("Final_S1andS2_Composite.tif")

## To check Sentinel-1 and Sentinel-2 combination image information#
#
S1and2_19may_image_SLL

## To change the band names##
names(S1and2_19may_image_SLL)<-c('Blue', 'Green', 'Red', 'RE1', 'RE2
',
                                 'RE3', 'NIR1', 'NIR2 or 8A', 'SWIR1', '
SWIR2', 'VV', 'VH', 'VH.VV')

## To make an image classification with the developed Random Forest
model##
classifiedS1andS2<-predict(S1and2_19may_image_SLL, LodgingscoreRFS1a
nd2_SLL, type='response', progress='window')

## To view the classified image informations##
classifiedS1andS2

par(mfrow=c(1,2))

## To plot the classified Sentinel-1 image##
plot(classifiedS1andS2)

## To export the classified image##
writeRaster(classifiedS1andS2, "D:/My MSc. Proposal and thesis
Documents/RS&FD/Sentinel1&Sentinel2/S1&2_to_be-
submmited/CSV/RF_S1andS2_classified_final/S1and2_May19_2018_RF_SLL_c
lassified_image_using_S2R_120samples.tif",
            format="GTiff", datatype='INT1U', overwrite=TRUE)
```