Agronomy Journal

ORIGINAL ARTICLE

Special Section: Machine Learning in Agriculture

A machine learning modeling framework for *Triticum turgidum* subsp. *durum* Desf. yield forecasting in Italy

Marco Fiorentini[1] | **Calogero Schillaci**[2] | **Michele Denora**[3] | **Stefano Zenobi**[1] | **Paola Deligios**[1,4] | **Roberto Orsini**[1] | **Rodolfo Santilocchi**[1] | **Michele Perniola**[3] | **Luca Montanarella**[2] | **Luigi Ledda**[1]

[1]Department of Agricultural, Food and Environmental Sciences (D3A), Agronomy and Crop Science, Marche Polytechnic University, Ancona, Italy

[2]European Commission, Joint Research Centre (JRC), Ispra, Italy

[3]Department of European and Mediterranean Cultures, Environment, and Cultural Heritage, University of Basilicata, Matera, Italy

[4]Department of Agricultural Sciences, University of Sassari, Sassari, Italy

Correspondence
Calogero Schillaci, European Commission, Joint Research Centre (JRC), Ispra, Italy.
Email: calogero.schillaci@ec.europa.eu

Assigned to Associate Editor David Clay.

Abstract

The forecasting of crop yield is one of the most critical research areas in crop science, which allows for the development of decision support systems, optimization of nitrogen fertilization, and food safety. Many tested modeling approaches can be differentiated according to the models and data used. The models used are traditional crop models that require data that are often difficult to measure. New modeling approaches based on artificial intelligence algorithms have proven to be of high performance, flexible, and can be tested based on available data. In this study, four independent field experiments conducted on *Triticum turgidum* subsp. *durum* Desf. in central–southern Italy were used to train a set of machine learning (ML) algorithms to predict the yield using 16 variables: fertilization, nitrogen management, pedoclimatic, and remote sensing data. Four ML algorithms were calibrated and validated over two independent sites, and a linear regression model was used as a control. The calibrated models can predict the grain yield in the two regions by using ancillary data, topsoil physical and chemical properties, multispectral drone imagery, climatic data, and nitrogen fertilizer applied at the site. Among the four ML algorithms, stochastic gradient boosting (root-mean-square error $= 0.58$ t ha^{-1}) outperformed others during calibration and transferability. Nitrogen application rate, seasonal precipitation, and temperature are the most important features for predicting wheat yield.

Abbreviations: CAP, common agricultural policy; GBM, gradient boosting machines; KNN, K nearest neighbor; LTE, long term experiment; MAE, mean absolute error; ML, machine learning; NDRE, normalized difference red edge index; NDVI, normalized difference vegetation index; RF, random forests; SVM, support vector machine; UAV, unmanned aerial vehicle; VI, vegetation index.

- -

1 | INTRODUCTION

Durum wheat (*Triticum turgidum* subsp. *durum* Desf.) is a cereal crop used in the production of pasta. It represents approximately 5% of global wheat production in the area planted (IGC, 2020). Durum wheat is cultivated in Italy, the Mediterranean basin, the southwestern and northern United States, Canada, and several other countries. Over the past 20 years, Italy has been the major durum wheat European producer (International grain council, www.igc.int/en) sowing an average of 1,635,764.5 ha of durum wheat (Diotallevi et al., 2015) and producing 4.0 MT per year (Diotallevi et al., 2015).

The prediction of within-field yields is a challenge in many rainfed agricultural systems. The ability to forecast wheat yield can provide growers with invaluable information to make informed decisions regarding in-season investments and management decisions, such as the application of N fertilizer (Acutis et al., 2014; Padilla et al., 2020). The early prediction of wheat yield can be valuable for planning management actions to lower environmental pressure and better estimate market revenues. There are many methods for cereal yield modeling, mainly based on phenology, soil, and weather data, which also support management practices, such as tillage, weeding, and nitrogen fertilization. On the one hand, deterministic-based methods DSSAT (Jones et al., 2003), APSIM (Holzworth et al., 2014; Kheir et al., 2021), Century (Lugato et al., 2014), SALUS (Basso et al., 2011), and ARMOSA (Valkama et al., 2020) allow for the simulation of each compartment, soil, plant, weather, and their fluxes. There are some drawbacks for applying these methods, such as the massive amount of data required (Mereu et al., 2021a), need for site-specific calibration (Hoogenboom et al., 2004), and spatial components (Basso et al., 2001).

On the other hand, stochastic methods based on field and remote sensing offer a new way to find a relationship between the biotic and environmental predictors and can be deployed in vast areas (Adamchuk et al., 2004; Chlingaryan et al., 2018; Kayad et al., 2019).

Within-field wheat yield can vary due to several factors, such as soil properties, occurrence of pests and diseases, and nitrogen management (Maestrini & Basso, 2018). Variability within the same field can be due to differences in management actions, for example, the amount of fertilizer, crop rotation, and variety (Colecchia et al., 2015). Since 1993, scientific research has contributed to more effective and efficient management of the common agricultural policy (CAP) through a broader range of technical support services to the Directorate-General Agriculture and Member-State Administrations (van der Velde & Nisini, 2019). Services have been developed to support European Union global agricultural monitoring and food security assessments. Agricultural monitoring is a crucial part of the integrated agricultural control system (Sagris et al., 2013), which is at the core of the CAP

Core Ideas

- Scalable machine learning algorithms were calibrated and tested.
- Nitrogen fertilization, pedo-climatic data, and remote sensing data are used as predictors.
- Stochastic gradient boosting algorithm performed better than random forest.
- Nitrogen fertilization, precipitation, and temperature are the essential features.
- The stochastic gradient boosting (root-mean-square error = 0.54 t ha^{-1}) outperformed other algorithms.

---

implementation in Europe. Monitoring activities are based on expertise in crop modeling, agro-meteorology, sampling methods, environmental geospatial analysis, econometrics, and European and global statistical services. Monitoring activities often require the development of control systems, such as land parcel management and remote sensing checks, for efficient implementation of the CAP. Additional modeling exercises are helpful for testing particular conditions to support rural development in coping with climate change (Mereu et al., 2021) and searching for adaptation strategies for agricultural-related policies. Long-term field trials offer a valuable source of information for implementing in-season management such as N fertilizer applications. However, these are spatially restricted methods that do not account for pedo-climatic differences. To deepen our understanding of these aspects, multi-site long-term trials are complementary to draw a big picture of climate-soil variability for wheat yield forecasts. Combining proximal and remote sensing with other data sources as input for the plot-scale yield forecast model is a crucial point to improve yield prediction accuracy (Fiorentini et al., 2021). In this field of agronomy research, there has been an increase in the number of studies that have been recently published. Filippi et al. (2020a) have assessed the potential of large yield-mapping datasets and diverse spatial covariates to forecast mid- and late-season within-field cotton yield.

The use of machine learning (ML) approaches can capture nonlinear relationships and has shown a strong ability to predict crop yield (Paudel et al., 2022a; Shendryk et al., 2021) when compared with traditional linear approaches (Filippi et al., 2019a; Shahhosseini et al., 2020). Furthermore, among the widely used algorithms, random forests (RF) and gradient boosting machines (GBM) have shown high predictive ability and improved accuracies (Leo et al., 2021). Artificial neural networks are also widely used in predictive modeling and soil mapping (Schillaci et al., 2021). ML is limited by how
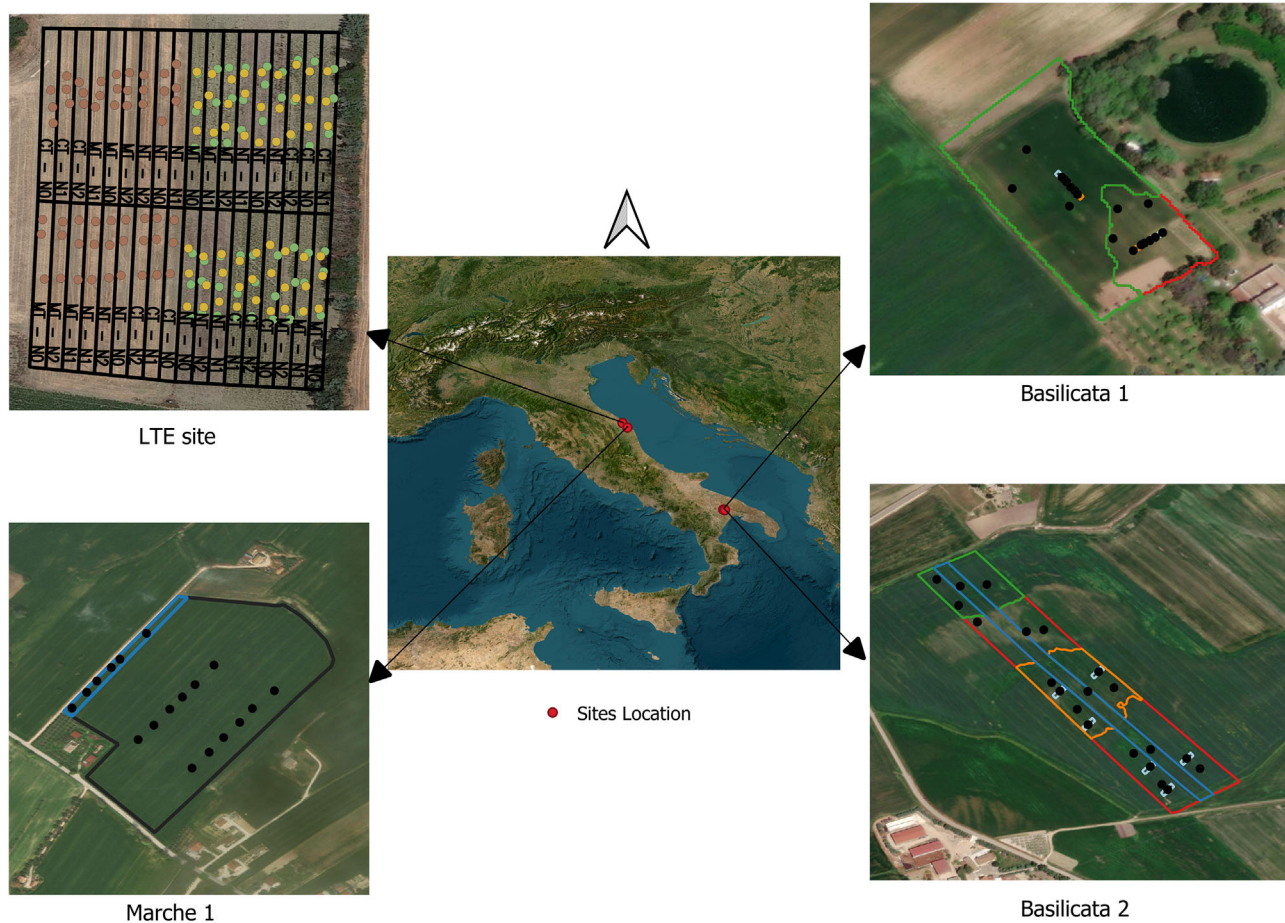
FIGURE 1 Location and experimental designs of the sites under analysis.

algorithms are trained with datasets that fail to completely represent possible combinations that can be found in the real world. When this occurs, we have a very unstable predictive performance (Chlingaryan et al., 2018).

For Italian durum wheat systems, similar investigations to evaluate the potential of ML algorithms with vegetation indices at different phenological stages, pedoclimate data, and N fertilization levels to predict within-field yield are limited.

This study aimed to develop a machine-learning framework for predicting the durum wheat yields based on topdressing nitrogen levels, pedoclimatic data, and remote sensing data. Soil organic carbon, bulk density, 11 different N fertilization levels, and vegetation indices acquired using an unmanned aerial vehicle (UAV) platform at four sites in central–southern Italy.

2 | MATERIALS AND METHODS

2.1 | Experimental site's overview

The study contained four sites located in the Marche region and two in the Basilicata region (Figure 1). All sites had different experimental designs and durum wheat (*Triticum*

turgidum subsp. *durum* Desf.) were grown for several years (Table 1).

The climate of the Marche experimental sites is meso-Mediterranean, whereas that of the two sites in Basilicata are Mediterranean based on the classification provided by the Walter and Leith climate diagram (Figure 2). Historical and in-season weather data are reported in Table 2.

All agronomic management activities performed at the four experimental sites during the growing season are shown in Table 2.

2.2 | Long-term experiment site

The long-term experimental (LTE) site was located in the "Pasquale Rosati" experimental farm of the Polytechnic University of Marche in Agugliano, Italy (43°32′ N, 13°22′ E, 100 m asl). This site is characterized by a mean annual precipitation of approximately 749 mm and a mean annual temperature of 15.5°C with monthly means ranging from 6°C in February to 26°C in July.

During the 2017–2018 crop growing season, the highest cumulative precipitation (647 mm) was recorded, whereas the 2018–2019 and 2019–2020 growing seasons recorded a

TABLE 1 Historical and in-season weather data of the four experimental sites.

Years	November	December	January	February	March	April	May	June	July
LET site									
Cumulative rainfall (mm)									
2017–2018	24	96	29	173	143	37	95	48	2
2018–2019	0	61	70	22	36	59	165	1	0
2019–2020	0	0	4	17	158	119	142	61	5
Long-term data 2000–2020	78	74	55	60	69	54	55	53	37
Mean temperature (°C)									
2017–2018	10	7	9	5	9	16	19	23	26
2018–2019	5	7	5	8	12	13	15	25	28
2019–2020	13	6	7	11	10	14	19	22	25
Long-term data 2000–2020	12	6	7	6	10	12	19	23	26
Marche 1									
Cumulative rainfall (mm)									
2019–2020	13	64	30	26	128	44	44	50	
Long-term data 2000–2020	96	84	50	48	51	65	51	47	
Mean temperature (°C)									
2019–2020	12.6	9	6.5	10.1	10.1	13.5	18.5	21.8	
Long-term data 2000–2020	12.3	9.2	7.9	7.7	10	12.8	17.4	21.7	
Basilicata 1									
Cumulative rainfall (mm)									
2020–2021			44	24	31	31	1	1	1
Long-term data 2000–2020			64	54	61	50	38	33	33
Mean temperature (°C)									
2020–2021			5.8	7.7	7.3	9.9	16	23.1	26
Long-term data 2000–2020			8.6	8.8	10.9	13.7	18	23.4	26
Basilicata 2									
Cumulative rainfall (mm)									
2019–2020			79	69	44	31	140	46	
Long-term data 2000–2020			64	54	61	50	38	33	
Mean temperature (°C)									
2019–2020			8	8.5	11.5	16.1	20	23.3	
Long-term data 2000–2020			8.6	8.8	10.9	13.7	18	23.4	

Abbreviation: LTE, long term experiment (1994–2022).

contraction of 36% (414 mm) and 22% (506 mm), respectively. The mean annual temperature in the 2019–2020 growing season was 1°C higher than the previous growing seasons (Table 1).

The soil of the study area has a silt-clay texture (Calcaric Gleyic Cambisols), 9.1 g kg^{-1} of organic carbon, 13.8 mg kg^{-1} of phosphorus (Olsen-P), 323.2 mg kg^{-1} of potassium, 26.7 cmol(+) kg^{-1} of cation exchange capacity, pH of 8.1, 29.6% of field water capacity, 17.6% of wilting point, and a slope of approximately 10%.

The LTE, established in 1994 (Orsini et al., 2019), consists of a rainfed 2-year rotation of durum wheat cultivar Tyrex
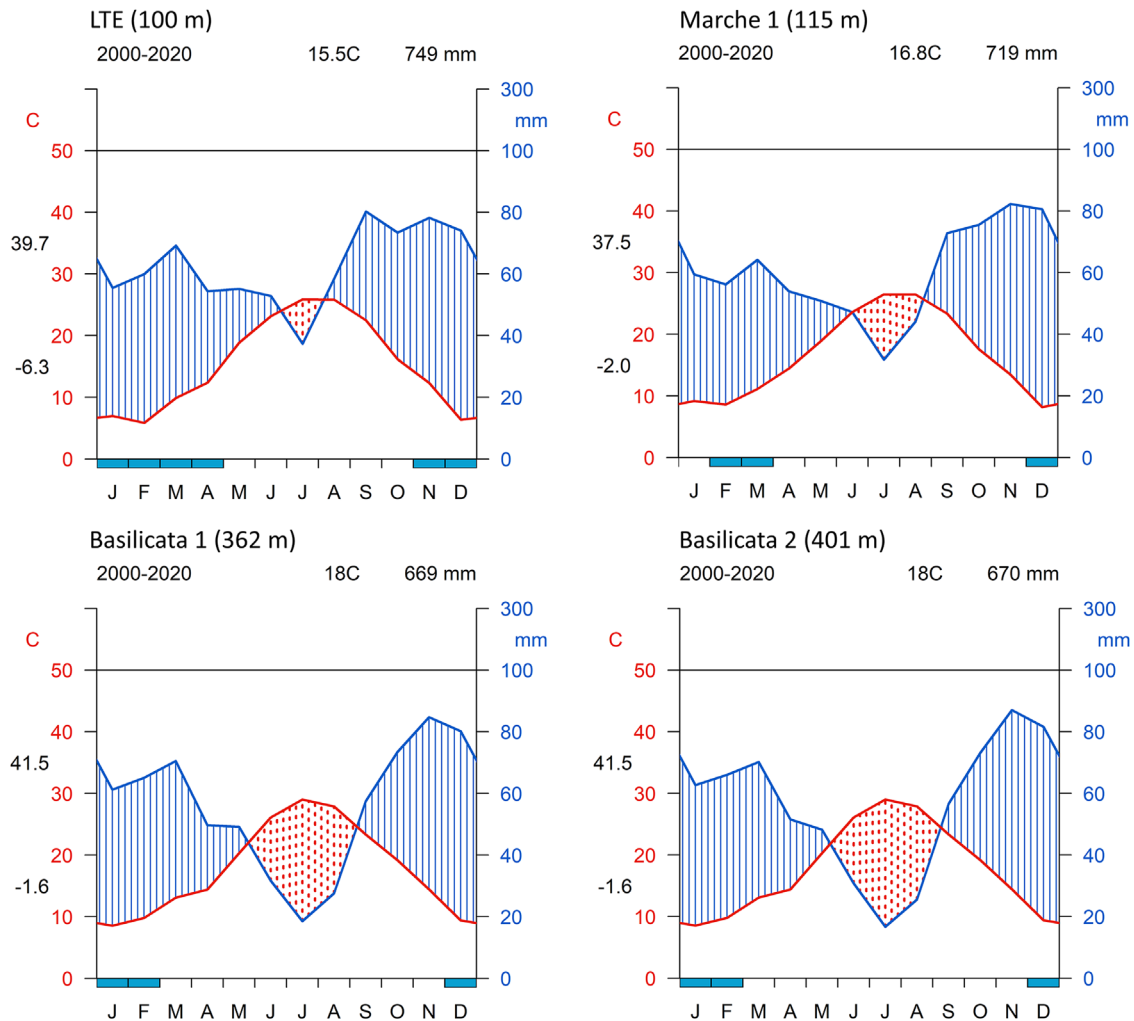
FIGURE 2 Walter and Lieth climate diagram of the four experimental sites (2000−2020 long-term series).

and maize (*Zea mays* L.) cv. DK440 (hybrid, FAO Class 300, Dekalb). Crop rotation was duplicated in two adjacent fields to allow all crops to be present each year (Figure 1). Within each field, three soil management (main plot, 1500 m^2 each) systems (conventional tillage; minimum tillage; and no-tillage) and three N fertilizer (sub-plot, 500 m^2 each) levels were arranged according to a split-plot experimental design with two randomized blocks and were repeated in the same plots every year. Conventional tillage, which is representative of the business-as-usual tillage practice in the study area, and minimum tillage plots were plowed along the maximum slope every year by a mouldboard (with two plows) at a depth of 40 cm or a chisel at a depth of 25 cm in autumn for wheat and in summer for maize. The seedbed was prepared by double harrowing prior to sowing. Durum wheat was sown at the beginning of autumn at a seeding rate of 220 kg ha^{-1}, with a row distance of 0.17 m. Each year, N fertilization in the form of urea (46%) was split into two applications: 50% at the end of tillering and 50% before head emergence. Weeds, pests, and diseases were controlled chemically. The sequences

of the agronomic practices applied at LTE are reported in Table 2.

The no-tillage soil was left undisturbed, except for sod seeding, crop residuals, weed chopping, and total herbicide spraying before seeding. The three N fertilizer treatments were N0, N90, and N180, corresponding to 0, 90, and 180 kg N ha^{-1}, respectively, distributed in two rates for wheat. The unfertilized N0 treatment was used as the control. The N90 treatment was compliant with the agri-environmental measures adopted within rural development plans at the local scale (https://www.regione.marche.it/Regione-Utile/Agricoltura-Sviluppo-Rurale-e-Pesca/Produzione-Integrata#Tecniche-Agronomiche). The N180 treatment represented conventional management.

2.3 | Marche 1 experiment description

The Marche 1 experiment was conducted during the 2019–2020 growing season at the Guzzini private farm in Recanati

TABLE 2 Agronomic management activities at the four experimental sites over the growing seasons.

Agro-technique	2018	2019	2020	2021
LET site				
Ploughing	10/02/2017	09/26/2018	11/11/2019	
Harrowing and seedbed preparation	11/20/2017	11/01/2018	11/25/2019	
Sowing[a]	11/21/2017	11/30/2018	12/27/2019	
Weed control: Glyphosate	03/28/2018	03/08/2019	02/10/2020	
N fertilization	03/29/2018	03/18/2019	02/25/2020	
	04/30/2018	05/02/2019	04/09/2020	
Pest control: Azoxystrobin, cyproconazole	04/24/2018	04/22/2019	04/20/2020	
Harvest	07/06/2018	07/07/2019	07/06/2020	
Marche 1				
Ploughing			10/02/2019	
Harrowing and seedbed preparation			11/11/2019	
Sowing			11/15/2019	
Weed control: Glyphosate			04/22/2020	
N fertilization			01/20/2020	
			03/20/2020	
Pest control: Azoxystrobin, cyproconazole			04/22/2020	
Harvest			06/28/2020	
Basilicata 1				
Ploughing				28/09/2020
Harrowing and seedbed preparation				14/12/2020
Sowing				18/01/2021
Weed control: Glyphosate and metsulfuron-methyl				23/03/2021
N fertilization				14/12/2020
				20/04/2021
Pest control: Pyraclostrobin				23/03/2021
Harvest				19/07/2021
Basilicata 2				
Harrowing and seedbed preparation			16/12/2019	
Sowing			07/01/2020	
Weed control: Glyphosate[a] and Metsulfuron-methyl			15/03/2020	
N fertilization			16/12/2019	
			14/04/2020	
Pest control: Pyraclostrobin			15/03/2020	
Harvest			20/07/2020	

Abbreviation: LTE, long-term experiment.

[a]Seed rate: 220 kg ha^{-1}, row spacing: 0.17 m.

(43°23′ N, 13°31′ E, 115 m asl), central Italy. The mean annual precipitation is 719 mm and the mean annual temperature is 16.8°C, with monthly means ranging from 7.7°C in February to 24.7°C in August. During the 2019–2020 crop growing season, the farm site was characterized by a seasonal rainfall of approximately 400 mm (approximately 100 mm lower than the precipitation recorded at Agugliano weather station during the same growing season) and by mean temperatures 1.5°C lower than the temperature recorded at Agugliano (Table 1). The soil of the farm site has a silt-clay texture, Calcaric Gleyic Cambisols (Seddaiu et al., 2016), with 8.4 g kg^{-1} of organic carbon, 10.3 mg kg^{-1} of phosphorus, 283.2 mg kg^{-1} of

potassium, 28.0 meq 100 g^{-1} of cation exchange capacity, pH of 8.1, 28.5% of field water capacity, 16.2% of wilting point, and a slope of about 6%.

The Marche 1 site was chosen because of the similar agronomic management, soil, and weather conditions. Indeed, the Marche 1 site had a rainfed 2-year crop rotation with durum wheat cv. Tyrex and maize cv. DK440 (Hybrid, FAO class 300). The soil was managed using CT, where two N levels were applied to two adjacent areas (Figure 1). Seedbed preparation was performed as reported at the LTE site, where plowing was performed along the maximum slope by a mouldboard plow (with two plows) at a depth of 0.4 m in autumn. Before seeding, a harrowing operation was performed while sowing was performed at a seed rate of 240 kg ha^{-1} with a row spacing of 0.13 m. The two N levels corresponded to 90 and 180 kg N ha^{-1} and were distributed at two rates. Each rate consisted of half of the total N and it was provided to the crop in the form of urea (46%). The first rate was provided at the end of tillering and the last rate before head emergence. Weeds, pests, and diseases were controlled chemically. The sequences of the agronomic practices applied at the farm site are shown in Table 2.

2.4 | Basilicata 1 experiment description

The Basilicata 1 experiment was conducted during the 2020–2021 growing season at the "Az. Agricola Pugliese" private farm located near Laterza (40°42′ N, 16°42′ E, 362 m asl), southern Italy. According to the Walter and Leith climatic classes, the climate of the study area (Figure 2c) was Mediterranean. The mean annual precipitation is 669 mm, and the mean annual temperature is 18°C, with monthly means ranging from 8.6°C in January to 26.4°C in July. During the 2020–2021 crop growing season, the farm site was characterized by a seasonal rainfall of 133 mm and a mean temperature of 13.7°C (Table 1). The soil of the Basilicata site 1 has a silt-clay texture (Haplic Calcisol) with 9.1 g kg^{-1} of organic carbon, 20.5 mg kg^{-1} of phosphorus (Olsen-P), 75.2 mg kg^{-1} of potassium, 17.88 cmol(+) kg^{-1} of cation exchange capacity, pH of 8.25, 33.4% of field water capacity, 20.6% of wilting point, and a slope of approximately 2%. The experiment was conducted using durum wheat (var. Kanakis, RAGT Semences, France).

Seedbed preparation was performed as reported in the LTE and Marche 1 sites, where plowing was performed along the maximum slope using a mouldboard plow (with two plows) at a depth of 0.30 m in autumn. Before seeding, a harrowing operation was performed in December, while sowing was done in January with a seed rate of 240 kg ha^{-1} with a row spacing of 0.13 m. Weed control was performed in March using glyphosate and metsulfuron-methyl, while Pyraclostrobin was used for pest control. An

experimental trial for precision agriculture was conducted during the 2020–2021 growing season. Variable-rate N fertilization was performed with a prescription map, as shown in Figure 1, where two homogenous areas were identified by the spatial variability of the soil with respect to physical and chemical characteristics (Denora et al., 2022).

The amount of N applied (urea 46%) was calculated based on the estimated N uptake by the crop and the soil characteristics. The N levels provided to the crop were 120 kg N ha^{-1} and 95 kg N ha^{-1}, split into two rates along the crop cycle, while within the two areas, we identified plots of 25 m^2 where the N levels applied were 150 kg N ha^{-1} and 0 kg N ha^{-1} (Figure 1). The sequences of the agronomic practices used at the farm site are shown in Table 2.

2.5 | Basilicata 2 experiment description

The Basilicata 2 experiments were carried out during the 2019–2020 growing season at the "Az. Agricola F. lli Lillo" private farm located in Matera (40°42′ N, 16°39′ E, 401 m asl), southern Italy. According to the Walter and Leith climatic classes, the climate of the study area (Figure 2d) was Mediterranean. The mean annual precipitation is 670 mm, and the mean annual temperature is 18°C, with monthly means ranging from 8.6°C in January to 23.4°C in June. During the 2019–2020 crop growing season, the farm site was characterized by a seasonal rainfall of 409 mm and a mean temperature of 14.5°C (Table 1). The soil of the Basilicata site 2 has a silt-clay texture (Haplic Vertisol) with 11.6 g kg^{-1} of organic carbon, 11.3 mg kg^{-1} of phosphorus (Olsen-P), 185.2 mg kg^{-1} of potassium, 21.2 cmol(+) kg^{-1} of cation exchange capacity, pH of 8.2, 32.2% of field water capacity, 18.9% of wilting point, and a slope of approximately 3%. The experiment was conducted using durum wheat (var. PR22D89, Pioneer Hi-Bred, Italy).

Seedbed preparation was performed as reported at the LTE, Marche 1, and Basilicata 1 sites, where plowing was performed along the maximum slope by a mouldboard plow (with two plows) at a depth of 0.30 m in autumn. Before seeding, a harrowing operation was performed in December, while sowing was done in January with a seed rate of 240 kg ha^{-1} with a row spacing of 0.13 m. Weed control was performed in March using glyphosate and metsulfuron-methyl, while Pyraclostrobin was used for pest control. An experimental trial for precision agriculture was conducted during the 2019–2020 growing season. Variable-rate nitrogen fertilization was performed with a prescription map, as shown in Figure 1, where two homogenous areas were identified by the spatial variability of the soil with respect to physical and chemical characteristics (Denora et al., 2022).

The amount of N applied (ammonium nitrate, 26%) was calculated based on the estimated nitrogen uptake by the crop and

TABLE 3 Experimental sites design and test areas description.

Site	Number of samples	Treatments	Area of samples (m^2)	Years of experiment	Total
LTE	18	N0	1	3 (2018, 2019, 2020)	162
	18	N90	1	3 (2018, 2019, 2020)	
	18	N180	1	3 (2018, 2019, 2020)	
Marche 1	6	N90	1	1 (2020)	18
	12	N222	1	1 (2020)	
Basilicata 1	6	N0	1	1 (2021)	18
	3	N95	1	1 (2021)	
	3	N120	1	1 (2021)	
	6	N150	1	1 (2021)	
Basilicata 2	6	N40	1	1 (2020)	21
	6	N118	1	1 (2020)	
	3	N133	1	1 (2020)	
	3	N139	1	1 (2020)	
	3	N150	1	1 (2020)	

Abbreviation: LTE, long-term experiment.

the soil characteristics. The N levels provided to the crop were 150, 139, 133, and 118 kg N ha^{-1}, whereas within the two areas, we identified plots of 4 m^2 where the N levels applied were 40 kg N ha^{-1} (Figure 1). The sequences of the agronomic practices used at the farm site are shown in Table 2.

2.6 | Proximal sensing measurements

To follow whole crop development, three phenological phases have been defined to perform UAV multispectral image acquisition. Multispectral images were acquired at crop tillering, Zadoks Scale 22 (ZS22; Zadoks et al., 1974), stem elongation, Zadoks Scale 35 (ZS35), and anthesis-Zadoks Scale 60 (ZS60) to compute the normalized difference vegetation index (NDVI) and normalized difference red edge index (NDRE). At crop maturity, Zadoks Scale 92 (ZS92), following the approach of Fiorentini et al. (2019). The test areas were randomly selected and georeferenced using the Leica Zeno 20 (Leica Geosystem) for each experimental site. Grain yield (t ha^{-1}) was measured for each test area. The grain yield (t ha^{-1}), expressed in dry matter, was measured using a laboratory thresher for each test area (1-m long row). The total number of test areas for each site is reported in Table 3.

2.7 | Unmanned aerial vehicle image acquisition and elaboration

Multispectral images were acquired at approximately 12:00 am using two different UAV platforms and multispectral cameras. For the Marche region sites, multispectral images were acquired using the MAIA S-2 multispectral camera (SAL Engineering) to compute NDVI and NDRE using the

following equations:

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \quad (1)$$

$$NDRE = \frac{(NIR - Red\ EDGE)}{(NIR + Red\ EDGE)} \quad (2)$$

For the Basilicata region sites, multispectral images were acquired using the Parrot Sequoia multispectral sensor (Parrot Drone SAS) to compute NDVI and NDRE. Both UAVs were equipped with an incident light sensor to avoid sun radiation variation during flight and perform better imagery acquisition, recording the light variation, and then correcting the spectral bands. Moreover, to obtain images with better spatial accuracy, ground control points were positioned in the area under investigation and then georeferenced before each flight operation. The relevant flight settings for each experimental site are listed in Table 4.

Each image acquired by a UAV flight requires an image processing operation for geometric correction, coregistration, and radiometric correction, conversion to the proper format file, and preparation of all image datasets for the orthomosaic map generation process. The image processing was composed of the following three main steps: (1) orthomosaic reflectance map generation, (2) computation of the vegetation indices (VIs) map, and (3) data extraction. After the flight, the raw images were exported from the multispectral cameras to the computer. Agisoft Metashape (Agisoft LLC) was used to generate the orthomosaic reflectance map. The software was based on the structure-from-motion algorithm (Verhoeven, 2011). To complete the second and third

TABLE 4 Unmanned aerial vehicle (UAV) flight setting information for Marche experimental sites.

Phenological stage	Date (mm/dd/yyyy)	Sensor	F-S overlap (%)	MOF	Resolution (cm² pixel⁻¹)	Flight time (min:s)	Flight speed (m s⁻¹)	Above ground level (m)
LTE								
ZS22	03/28/2018	MAIA S-2	75–65	Automatic	2.7	12:20	3.9	55
ZS35	04/28/2018	MAIA S-2	75–65	Automatic	2.7	12:20	3.9	55
ZS60	05/27/2018	MAIA S-2	75–65	Automatic	2.7	12:20	3.9	55
ZS22	03/15/2019	MAIA S-2	75–65	Automatic	2.7	12:20	3.9	55
ZS35	04/18/2019	MAIA S-2	75–65	Automatic	2.7	12:20	3.9	55
ZS60	05/24/2019	MAIA S-2	75–65	Automatic	2.7	12:20	3.9	55
ZS22	02/24/2020	MAIA S-2	75–65	Automatic	2.7	12:20	3.9	55
ZS35	02/23/2020	MAIA S-2	75–65	Automatic	2.7	12:20	3.9	55
ZS60	04/02/2020	MAIA S-2	75–65	Automatic	2.7	12:20	3.9	55
Marche 1								
ZS22	02/17/2020	MAIA S-2	75–65	Automatic	3.5	09:17	10.2	70
ZS35	03/19/2020	MAIA S-2	75–65	Automatic	3.5	09:17	10.2	70
ZS60	04/24/2020	MAIA S-2	75–65	Automatic	3.5	09:17	10.2	70
Basilicata 1								
ZS22	03/28/2021	Parrot Sequoia	75–65	Automatic	7.54	07:03	2	80
ZS35	04/28/2021	Parrot Sequoia	75–65	Automatic	7.54	07:03	2	80
ZS60	05/27/2021	Parrot Sequoia	75–65	Automatic	7.54	07:03	2	80
Basilicata 2								
ZS22	03/15/2020	Parrot Sequoia	75–65	Automatic	8.48	11:50	6	90
ZS35	04/18/2020	Parrot Sequoia	75–65	Automatic	8.48	11:50	6	90
ZS60	05/24/2020	Parrot Sequoia	75–65	Automatic	8.48	11:50	6	90

Abbreviations: F-S, forward-side overlap; LTE, long-term experiment; MOF, mode of flight; UAV, unmanned aerial vehicle; ZS, Zadoks scale.

main steps, the orthomosaic reflectance map was imported into R statistical software (Core Team, 2014), and then the "stack" function of the raster R package (Hijmans et al., 2011) was used. Based on previous studies (Fiorentini et al., 2021; Orsini et al., 2019a, 2020; Pro et al., 2021), NDVI and NDRE are VIs with a high correlation with durum wheat crop parameters and are computed using Equations (1) and (2), where Red (665 nm) is the red band, Red Edge is the red-edge spectral band (705 nm), and NIR is the near-infrared spectral band (865 nm). To extract the NDVI and NDRE values only at the points where the sampling was obtained, we imported a .xlsx file containing the georeferenced positions of samples by using the "read_excel" function of the *readxl* R package (Wickham, 2016). Then, the "extract" function of the *raster* R package (Hijmans et al., 2011) was used to extract the values in a data frame format and the created data-frame object was exported in a CSV file format using the "write.csv" function of the *utilis* R package (Wickham et al., 2019).

2.8 | Soil data

Soil organic matter was determined according to the modified Walkley–Black method (Walkley & Black, 1934); the Van Bemmelen factor (= 1.724) has been used to obtain soil organic carbon; soil texture according to the pipette method (Gee & Bauder, 1986); and available phosphorous (Olsen, 1954), potassium (Della Chiesa et al., 2019), and cation exchange capacity (Rhoades, 1983). The ratio of carbon to nitrogen (C/N), organic matter, and nitrogen (N) content were measured for each experimental site. Thirty six samples were acquired at the LTE site, six soil samples were acquired at the Marche 1 site, six soil samples were acquired at Basilicata 1, and six soil samples were acquired at Basilicata 2. All soil samples were acquired before sowing at each site under analysis. Only soil organic carbon, N, and C/N ratio have been used as variables.

2.9 | Modelling approach

The data provided to the different ML algorithms are comprehensive in fertilization management, climatic data such as temperature and precipitation, and chemical properties of the soil and VIs. This allows us to provide all the available data to the different ML algorithms to better describe the weather–soil–crop system (Figure 3).

The fertilization management input is related to the N input (kg N ha^{-1}) provided to the crop for all sites, and all fertilization management information is reported in Table 3. Climatic data such as the mean temperature (°C) and the sum of precipitation (mm) were calculated from the UAV flight performed

at ZS22 to the UAV flight performed at ZS60 by using the free accessible climate data NASA Prediction of Worldwide Energy Resources (POWER) using the NASA POWER R package (Sparks, 2018).

The NASA POWER climate data are used in many deterministic models. Moreover, this allows us to scale the use of this developed model to the regional level because NASA POWER covers the entire globe.

The NDVI and NDRE (Fiorentini et al., 2021; Orsini et al., 2019, 2020; Pro et al., 2021) can be calculated from satellite images, such as the Copernicus Satellite System, which covers the entire globe or commercial satellite systems.

To provide a temporal component and thus the crop growth development to the ML algorithms, the difference in the values of the NDVI and NDRE between the three UAV flights (ZS22, ZS35, and ZS60) was calculated using the following equations:

$$\Delta \text{ NDRE } 35 \ 22 \ = \ \text{NDRE ZS35} \ - \ \text{NDRE ZS22} \quad (3)$$

$$\Delta \text{ NDRE } 60 \ 35 \ = \ \text{NDRE ZS60} \ - \ \text{NDRE ZS35} \quad (4)$$

$$\Delta \text{ NDRE } 60 \ 22 \ = \ \text{NDRE ZS60} \ - \ \text{NDRE ZS22} \quad (5)$$

$$\Delta \text{ NDVI } 35 \ 22 \ = \ \text{NDVI ZS35} \ - \ \text{NDVI ZS22} \quad (6)$$

$$\Delta \text{ NDVI } 60 \ 35 \ = \ \text{NDVI ZS60} \ - \ \text{NDVI ZS35} \quad (7)$$

$$\Delta \text{ NDVI } 60 \ 22 \ = \ \text{NDVI ZS60} \ - \ \text{NDVI ZS22} \quad (8)$$

2.10 | Machine learning algorithms

Five different ML methods were compared: (1) linear model (LM) as the benchmark algorithm, (2) support vector machine (SVM; Cortes & Vapnik, 1995), (3) K nearest neighbors (KNN; Dasarathy, 1990), (4) RF (Svetnik et al., 2003), and (5) stochastic gradient boosting (Friedman, 2002). These ML algorithms were chosen because they can be easily compared using the scientific literature (Chlingaryan et al., 2018). An SVM has the potential to resolve the problem of overfitting when analyzing high-dimensional data to solve both regression (Chlingaryan et al., 2018) and classification tasks (Yang et al., 2016), such as the prediction of nitrogen content and yield or image classification crop-weed. The SVM algorithm was optimized by adjusting the regularization constant, and the best model was selected based on the lowest root-mean-square error (RMSE).
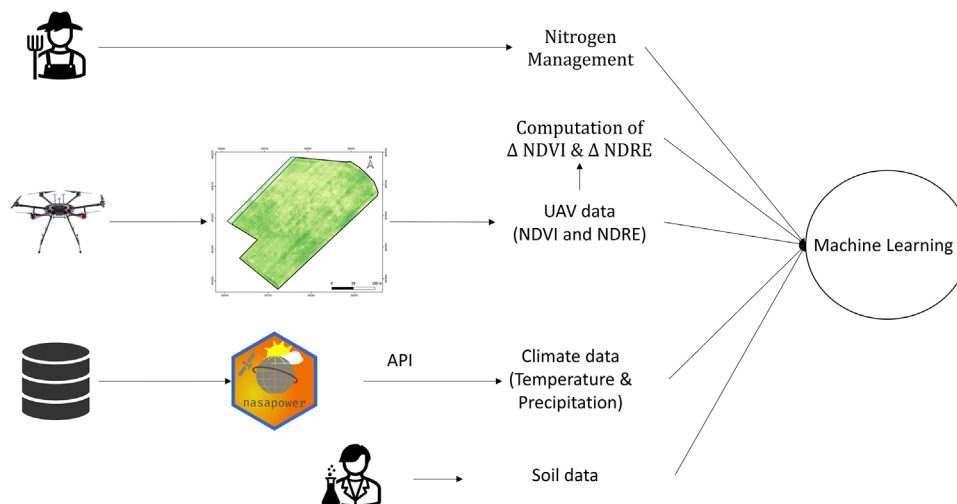
FIGURE 3 Source data and machine learning approach. NDRE, normalized difference red edge index; NDVI, normalized difference vegetative index; UAV, unmanned aerial vehicle.

TABLE 5 Algorithm name, abbreviation, R package used, and list of the hyperparameters optimized.

Algorithm	Abbreviation	R package	Tuning hyperparameter
Linear model	LM	base	Intercept
Support vector machine	SVM	kernelab	regularization constant
K nearest neighbors	KNN	kknn	kmax, distance, kernel
Random forest	RF	random forest	Mtry
Stochastic gradient boosting	GBM	gbm	ntrees, interaction.depth, shrinkage, n.minobsinnode

TABLE 6 Summary statistics of the measured grain yield (t ha^{-1}) of the four experimental sites.

Site	Mean	Standard deviation	Maximum	Minimum	Interquartile range
LTE	3.75	1.72	6.85	0.71	2.67
Marche 1	5.5	0.73	6.52	4.02	1.25
Basilicata 1	3.62	1.29	4.87	1.44	2.2
Basilicata 2	5.18	0.88	6.67	3.45	0.89

Note: Model calibration and transferability.

Abbreviation: LTE, long-term experiment.

The KNN is an intuitive algorithm that classifies or predicts unlabeled instances using examples with known labels (L. Zhang et al., 2010). KNN attempts to find the k closest instances in the training set and assigns the instance to the label that appears most frequently within the k-subset, which can be done by taking votes if the values of the target function are discrete or by computing the mean if the values are continuous (Mitchell, 1997).

RF use an ensemble-learning method consisting of many decisions or regression trees and (Breiman, 2001; Kim & Lee, 2016) a bagging technique to split the dataset into homogenous subsets (trees) in parallel. When building each tree, RF randomly samples the training data, and a random subset of features is used to create a predictive model. Final predictions are made by combining (bagging) all trees/models and using the average predicted results.

Similar to RF, stochastic gradient boosting uses an ensemble learning method but uses boosting rather than bagging. Model predictions are improved by sequentially converting weak learners (variables) that are poorly correlated with the target variable into strong and well-correlated learners (Friedman, 2002a). The GBM algorithm was tuned by adjusting the maximum tree depth (interaction depth), number of boosting iterations or trees (ntrees), shrinkage, and minimum terminal node size (n.minobsinnode), and the model with the lowest RMSE was selected.

All ML model hyperparameters were adjusted using a 15-grid search and fivefold cross-validation. The models were evaluated against observed yields using the coefficient of determination (R^2), RMSE, and mean absolute error (MAE).

Moreover, the *varImp* function of the caret R package was used to visualize the variable importance to describe which covariate contributed the most to improve the accuracy of the models (Kuhn, 2008). The permutation-based variable importance was used (Altmann et al., 2010), which offers several advantages. It is a model-agnostic approach for assessing the influence of an explanatory variable on a model's performance. The plots of the variable importance measures are easy to understand because they are compact and present essential variables in a single graph. These measurements can be compared between models and may lead to interesting insights.

By contrast, in regularized regression models, the effect of one variable may dominate the impact of other correlated variables. The main disadvantage of the permutation-based variable importance measure is its dependence on the random nature of the permutations. Consequently, different permutations generally produce different results. Moreover, the value of the measure depends on the choice of loss function.

To compare the performances of the algorithms, the caret R package (Kuhn, 2008) was adopted. This allows for the use of different algorithms in the same modeling framework. The R package used to create the models is reported in Table 5.

The algorithms were trained based on the LTE and Basilicata 1 experimental sites with 183 data points. In contrast, the Marche 1 and Basilicata 2 experimental sites were chosen as test datasets to measure the accuracy of the models with 36 data points.

3 | RESULTS

3.1 | Summary statistics

The results show that the grain yield (t ha^{-1}) obtained from the four sites is different, with a mean value of 3.75 t ha^{-1} for the LTE site, 5.50 t ha^{-1} for the Marche 1 site, 3.62 t ha^{-1} for Basilicata 1 site, and 5.18 t ha^{-1} for the Basilicata 2 site (Table 6). Instead of considering the interquartile range, we had a higher interquartile range for Agugliano (2.67) and Basilicata 1 (2.20) and a lower interquartile range Marche 1 (1.25) and Basilicata 2 (0.89).

3.2 | Model calibration and transferability

The results of the fivefold cross-validation are presented in numerical format in Table 7.

The GBM and RF models obtained significantly lower RMSE and MAE values and considerably higher R^2 values compared with the LM, SVM, and KNN models.

TABLE 7 Machine learning algorithms, abbreviation, root mean square error (RMSE), mean absolute error (MAE), and r squared value.

Model	Abbreviation	RMSE	MAE	R^2
Linear model	LM	0.68	0.55	0.85
Support vector machine	SVM	0.69	0.55	0.85
K nearest neighbors	KNN	0.65	0.47	0.86
Random forest	RF	0.47	0.37	0.93
Stochastic gradient boosting	GBM	0.48	0.37	0.92

The results demonstrate that the GBM and RF models performed better than the other models evaluated in this work, with RMSE values of 0.48 t ha^{-1} and 0.47 t ha^{-1}, respectively, whereas the LM, SVM, and KNN models obtained RMSE values of 0.68, 0.69 and 0.65 t ha^{-1}, respectively. The hyperparameter for each model is listed in Table 8.

Model transferability was tested using a dataset that did not include the models during the training procedure (Chollet & Allaire, 2018). Table 9 lists the transferability error performance of the model.

The results demonstrate that the GBM model performed better than the other models (Table 9) as it obtained lower RMSE and MAE values and considerably higher R^2 values compared with the other models.

GBM obtained lower RMSE values of 0.3, 1.02, 0.5, and 0.41 for RF, KNN, SVM, and LM, respectively. In comparison, the worst model in terms of metric transferability errors was KNN, which had an RMSE of 1.60, which was unacceptable considering the interquartile range of the Marche 1 and Basilicata 2 datasets. The GBM also had a lower confidence interval than the other models (Figure 4).

3.3 | Feature importance

The variable importance of all the models is shown in Figures 5–7. The nitrogen variable was the most important feature in all models except the linear model. Based on of the feature importance of the ML models, the three most important variables in descending order were nitrogen, precipitation, and temperature. While considering the feature importance of the best model, such as the GBM, the five most important features are nitrogen management, precipitation, temperature, nitrogen soil content, and NDVI ZS22.

4 | DISCUSSION

4.1 | Model calibration and transferability

In this study, we compared the capability of predicting durum wheat yield in central–south Italy using several ML algorithms that incorporate nitrogen fertilization level,

TABLE 8 Hyperparameter definition for the model calibration procedure.

Model	Abbreviation	Tuning hyperparameter
Linear model	LM	Intercept = True
Support vector machine	SVM	Regularitation constant = 1
K nearest neighbors	KNN	Kmax = 5, distance = 2, kernel = optimal
Random forest	RF	mtry = 31
Stochastic gradient boosting	GBM	n. trees = 100, interaction depth = 3, shrinkage = 0.1, n. minobsinnode = 10

TABLE 9 Transferability models error metric.

Model	Abbreviation	RMSE	MAE	R^2
Linear model	LM	0.99	0.81	0.56
Support vector machine	SVM	1.08	0.94	0.43
K nearest neighbors	KNN	1.60	1.14	0.03
Random forest	RF	0.81	0.66	0.71
Stochastic gradient boosting	GBM	0.58	0.48	0.84

Abbreviation: MAE, mean absolute error; RMSE, root mean square error.

pedoclimatic, and remote sensing data as source data. Previous studies have primarily focused on the use of on-farm or remote-sensing data sources (Feng et al., 2020; Haghverdi et al., 2018; Huang et al., 2016; Meng et al., 2019).

In Greece, Dietrich et al. (2022) used two modeling approaches to predict durum wheat yield. The first approach used a multiple linear regression model trained with the enhanced vegetation index (EVI) and the normalized multi-band drought index, and the second used the sentinel 2 bands for several dates as input of three ML algorithms: RF, KNN, and GBM. The RF and KNN models achieved higher accuracy with $R^2 > 0.87$ and RMSE < 455 kg ha^{-1}. In this study, data on pedoclimate and fertilization levels were provided to the algorithms to be trained. In Mexico (Gómez et al., 2021), the performance of eight different algorithms to predict durum yield with climate and remote sensing data was evaluated using feature and no-feature selection methods. The authors showed that RF with no feature selection achieved an R^2 of 0.84.

Other studies have focused on publicly available datasets but were either conducted on one field site or neglected the squaring growth stage when N fertilizer is commonly applied (Filippi et al., 2020; Nguyen et al., 2019).

In the current study, a combination of different sources improved the model's performance. These findings agree with other studies, where more data improved yield predictions (Filippi et al., 2019; 2020; Nguyen et al., 2019), particularly with Paudel et al. (2022), which reported a median normalized RMSE of 7.49% for soft wheat.

Moreover, the in situ remote sensing used in this study could be replaced by free multispectral satellite data, such as Sentinel 2.

During the calibration model procedure that used fivefold cross-validation, the GBM and RF models were found to be the best models (Table 7). This is in agreement with other studies predicting yield or other crop parameters (Kayad et al., 2019; Liang et al., 2015; Richetti et al., 2018; Wu et al., 2019).

The superior performance of GBM and RF compared to the other models is attributed to how each ML algorithm builds the optimal model. Both GBM and RF use ensemble learning methods, in which several models are combined to improve the overall model performance (Zhang & Ma, 2012). In GBM, weak learners (variables) are sequentially converted to strong learners to decrease bias from highly correlated variables. In contrast, RF generates models (trees) from a random subset of the training data and averages all trees to reduce the variance (Leo et al., 2021).

To verify that the models we created are reliable, scalable, and usable, they were tested using data that models have not processed during the calibration procedure (Chollet & Allaire, 2018). Two datasets were provided to test the transferability of the model between the Marche region (Marche 1) and the Basilicata region (Basilicata 2). The GBM model obtained a statistically higher performance, with the lowest RMSE and MAE values and a higher R^2. Moreover, the GBM was the most stable model in terms of metric errors and obtained the lowest variation in error metrics, RMSE, MAE, and R^2, between calibration and transferability. The worst model was
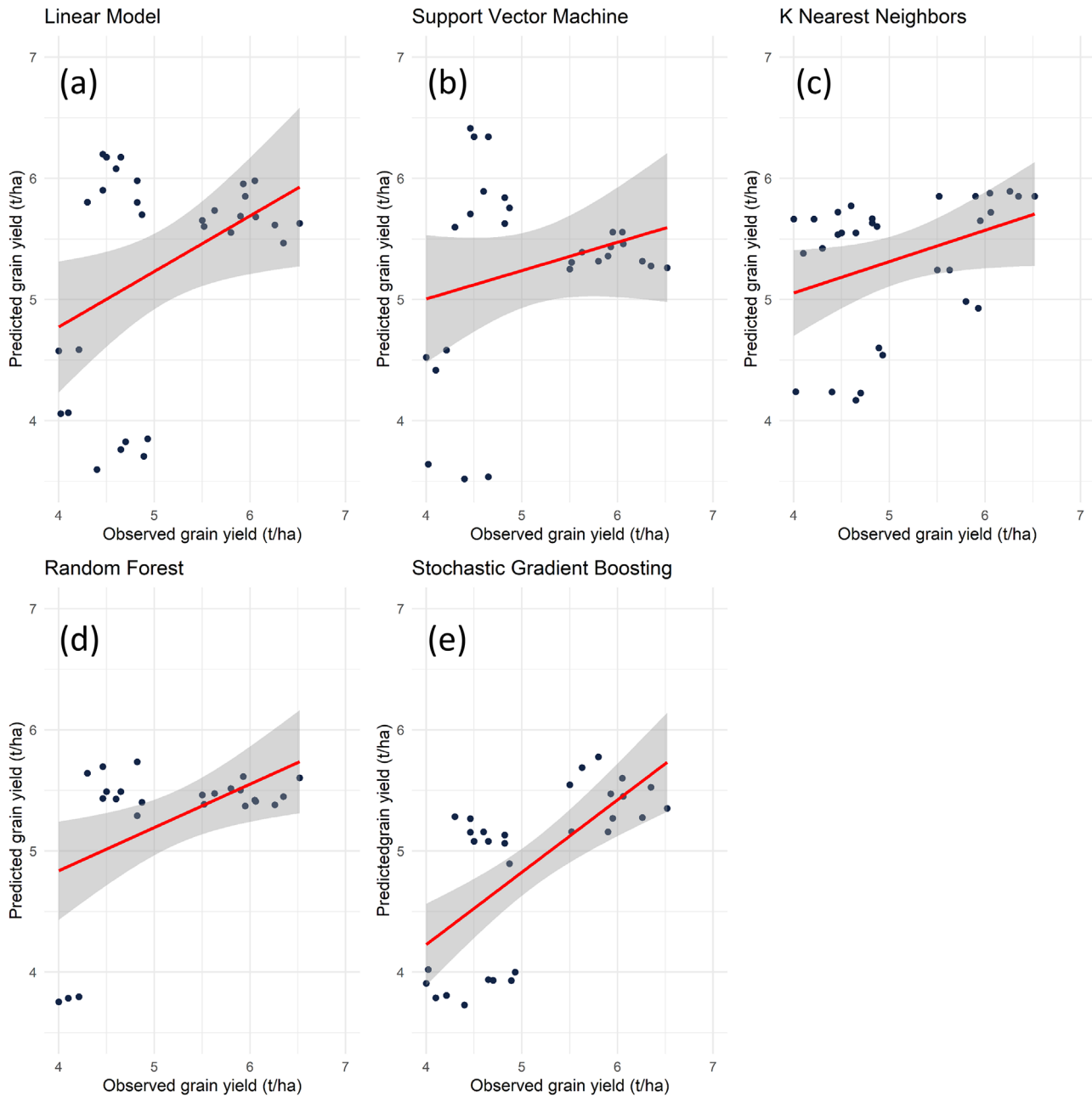
FIGURE 4 Observed versus predicted grain yield (t h^{-1}) for the (a) linear model, (b) support vector 384 machine, (c) k nearest neighbors, (d) random forest, and (e) stochastic gradient boosting.

LM, as observed in several studies, that used LM as a control algorithm (Chlingaryan et al., 2018).

4.2 | Feature importance

Variable importance is valuable for interpreting the models obtained and providing helpful information for future ML modeling approaches.

N was the most important feature for improving the prediction accuracy of all the models, except for linear regression.

This agrees with several studies that have revealed that nitrogen is a critical driver of durum wheat grain yield (Fiorentini et al., 2021 Orsini et al., 2020; Pro et al., 2021; Seddaiu et al., 2016).

Nayak et al. (2022) also reported that nitrogen was the covariate that improved the accuracy of the RF trained with a multidata source approach.

Mineral nitrogen fertilization is fundamental to increase the concentration of chlorophyll in leaves (Fiorentini et al., 2019) and metabolic processes that enable the production of

Feature Importance
created for the Liner Model, Support Vector Machine model
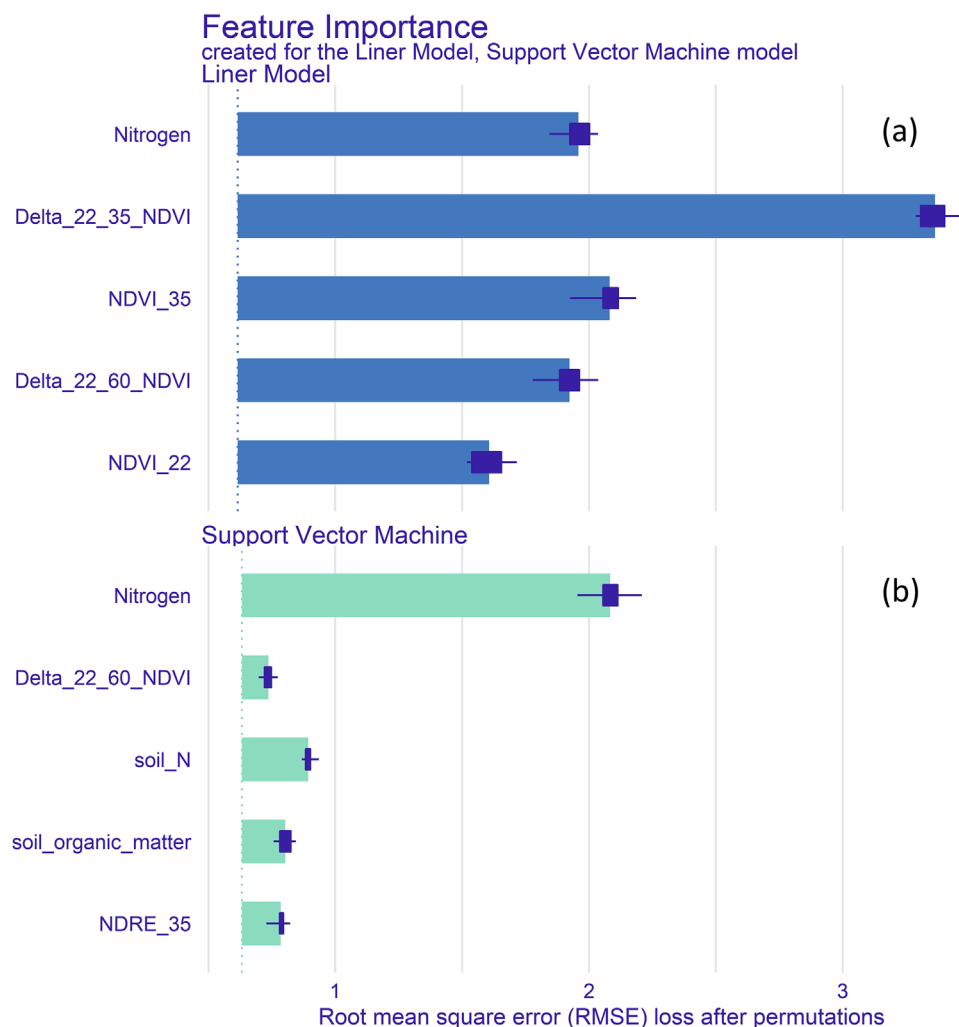
FIGURE 5 Variable importance for (a) linear regression (b) and support vector machine algorithms. NDRE, normalized difference red edge index; NDVI, normalized difference vegetation index.

amino acids, and thus, strongly affects grain quality (Abad et al., 2004).

Optimization of nitrogen supply is an important research topic for policy makers and farm decision makers, especially from an environmental perspective (Adenuga et al., 2019). After this abrupt geopolitical shift, nitrogen fertilizer prices have increased, and many countries are creating strategies to ensure that there is sufficient fertilizer available. The farm decision maker aims to reduce corporate production costs and reduce the farm carbon footprint (Menegat et al., 2022). From an environmental point of view, it is important to optimize nitrogen because excess can result in nitrate leaching (Rath et al., 2021).

To compare different nitrogen fertilization management practices in different crops, the agronomic research community employs nitrogen use efficiency, which is a dimensionless value used to calculate the nitrogen uptake from the crop (Rütting et al., 2018) and compare different nitrogen management practices.

In addition to N, other important variables for improving the accuracy of the model were the mean temperature in January and the amount of precipitation in April, while no information regarding the vegetation indices was reported. This is in accordance with the reports of Seddaiu et al. (2016) who studied the factors influencing N fertilization.

Precipitation is important not only for providing the water needed for crop growth but also for the nutrients in fertilizers to enter the circulating solution from which the crop is able to uptake agronomic input (Z. Li et al., 2022).

Temperature controls crop development; numerous studies have shown that degree-days can well model the entire phenological development of the crop (Q. Y. Li et al., 2012). Moreover, multi-source data and several ML algorithms have been used to predict wheat grain yields in China (Han et al., 2020). By calculating the decreased accuracy (mean square error) after removing one variable from the RF model, the authors showed the order of variable importance of the EVI, minimum temperature, precipitation, NDVI, and soil
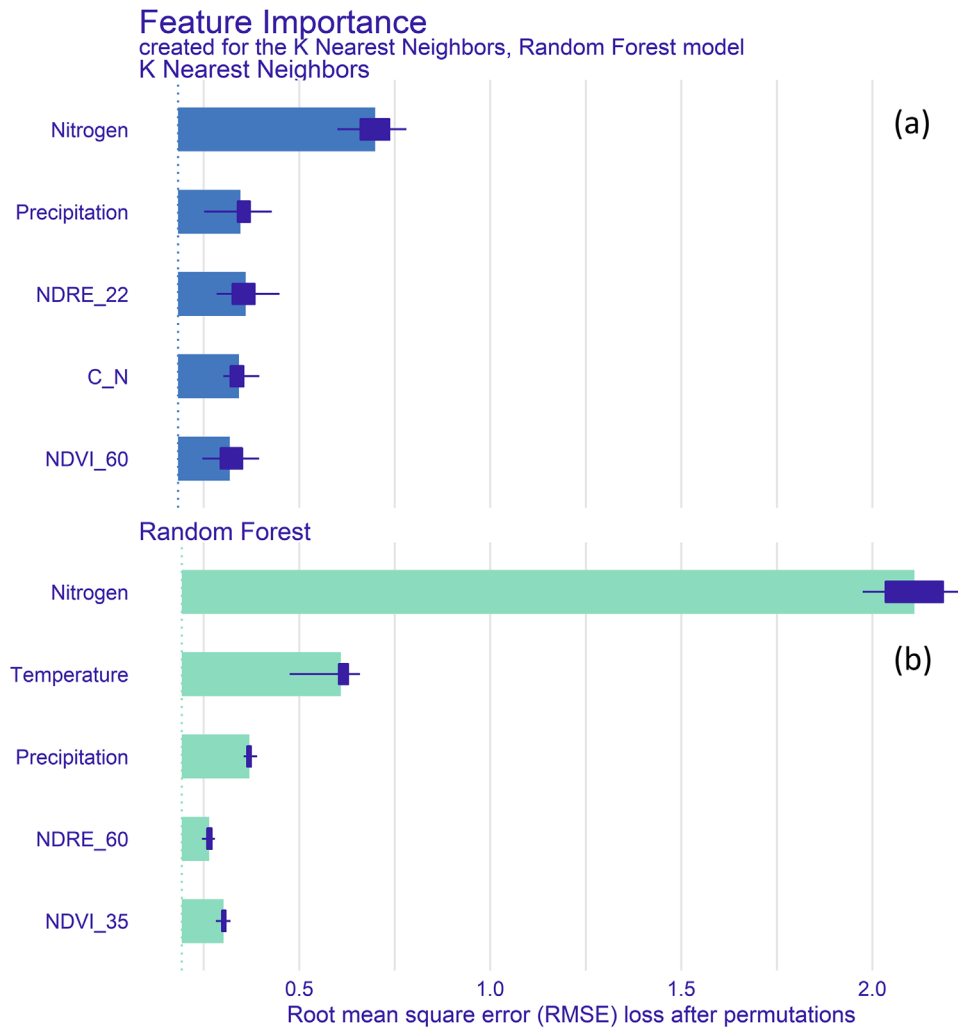
FIGURE 6 Variable importance for (a) k nearest neighbor (b) and random forest algorithms. NDRE, normalized difference red edge index; NDVI, normalized difference vegetation index.

moisture. The soil nitrogen content led to a higher accuracy for the GBM model.

5 | CONCLUSIONS

One of the most important research topics in agriculture and food safety is the development of a model that can predict high-accuracy on-season durum wheat yield. It is crucial to consider covariates that can represent the climate, soil, crop, and agronomic management systems.

In this study, we explore the possibility of combining several different source data, such as nitrogen management, pedoclimatic, and remote sensing data of four Italian experimental sites, to train and test five ML algorithms to predict durum wheat yield.

Stochastic gradient boosting was the best ML algorithm, with a 0.58 RMSE of the test set and a lower error metric variation between calibration and transferability.

It was observed that N, precipitation, and temperature were the features that helped improve the model's accuracy. Therefore, they must be considered in future grain yield ML modeling exercises.

The generated model can be scaled for the Marche and Basilicata regions using the soil samples acquired in the field, the multispectral satellite images from the Copernicus program, and the meteorological data from NASA.

AUTHOR CONTRIBUTIONS

Marco Fiorentini: Conceptualization; data curation; formal analysis; methodology; software; validation; writing—original draft. **Calogero Schillaci**: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; software; supervision; validation; visualization; writing—original draft; writing—review and editing. **Michele Denora**: Data curation; investigation; validation; visualization. **Stefano Zenobi**: Data curation; methodology; validation; visualization. **Paola Deligios**: Data
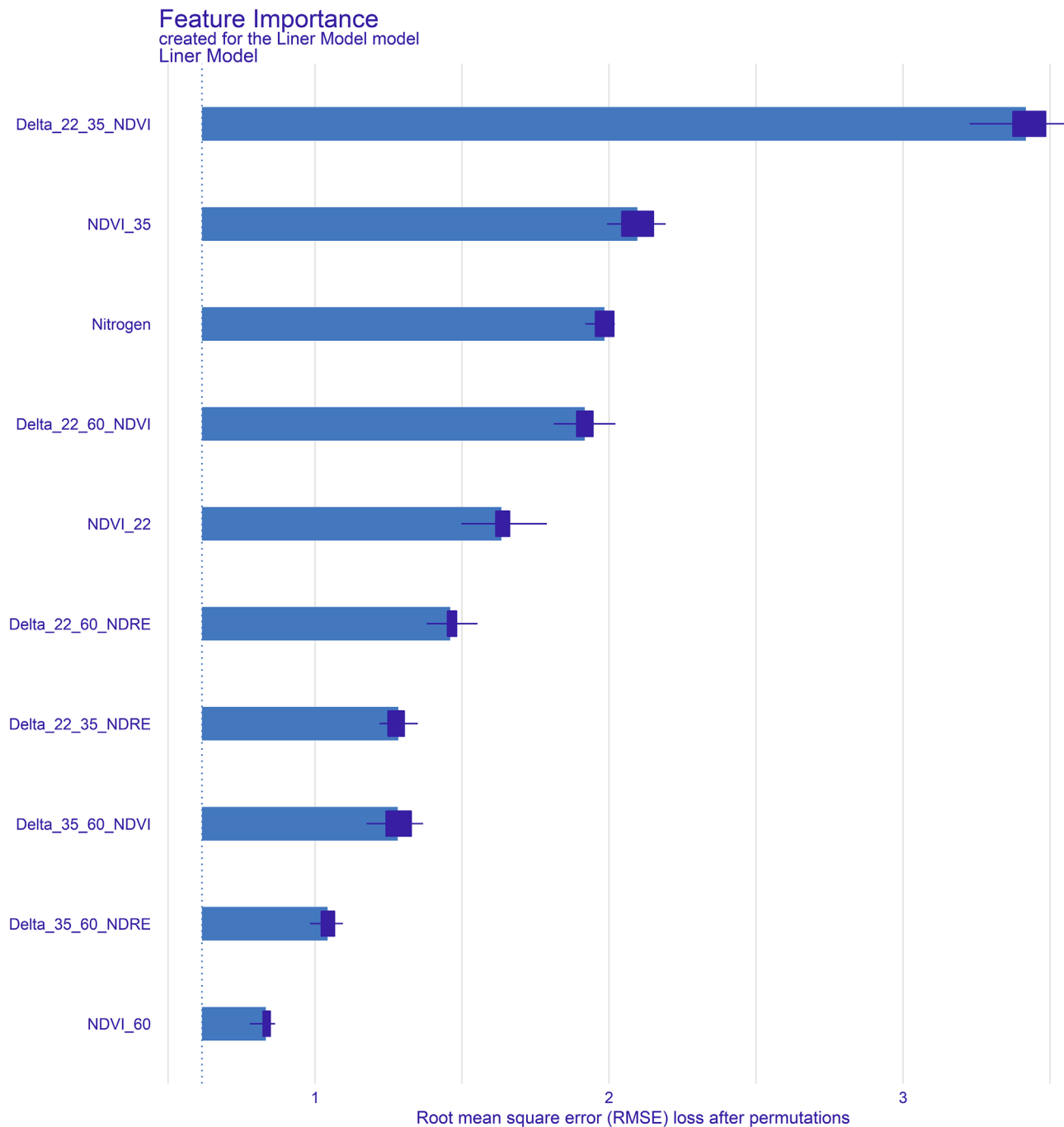
FIGURE 7 Variable importance for the stochastic gradient boosting algorithm. NDRE, normalized difference red edge index; NDVI, normalized difference vegetation index.

curation; formal analysis; investigation; methodology; supervision; writing—review and editing. **Roberto Orsini**: Data curation; supervision. **Rodolfo Santilocchi**: Methodology; validation. **Michele Perniola**: Data curation; investigation; resources; supervision; writing—review and editing. **Luca Montanarella**: Methodology; supervision; writing—review and editing. **Luigi Ledda**: Conceptualization; data curation; formal analysis; funding acquisition; investigation;

methodology; project administration; resources; supervision; validation; writing—review and editing.

CONFLICT OF INTEREST STATEMENT

ORCID

Marco Fiorentini https://orcid.org/0000-0002-1101-7313
Calogero Schillaci https://orcid.org/0000-0001-7689-5697
Paola Deligios https://orcid.org/0000-0001-9724-2812

REFERENCES

Abad, A., Lloveras, J., & Michelena, A. (2004). Nitrogen fertilization and foliar urea effects on durum wheat yield and quality and on residual soil nitrate in irrigated Mediterranean conditions. *Field Crops Research*, *87*(2–3), 257–269. https://doi.org/10.1016/J.FCR.2003.11.007

Acutis, M., Alfieri, L., Giussani, A., Provolo, G., Guardo, A. D., Colombini, S., Bertoncini, G., Castelnuovo, M., Sali, G., Moschini, M., Sanna, M., Perego, A., Carozzi, M., Chiodini, M. E., & Fumagalli, M. (2014). ValorE: An integrated and GIS-based decision support system for livestock manure management in the Lombardy region (northern Italy). *Land Use Policy*, *41*, 149–162. https://doi.org/10.1016/j.landusepol.2014.05.007

Adamchuk, V. I., Hummel, J. W., Morgan, M. T., & Upadhyaya, S. K. (2004). On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture*, *44*(1), 71–91. https://doi.org/10.1016/j.compag.2004.03.002

Adenuga, A. H., Davis, J., Hutchinson, G., Donnellan, T., & Patton, M. (2019). Environmental efficiency and pollution costs of nitrogen surplus in dairy farms: A parametric hyperbolic technology distance function approach. *Environmental and Resource Economics*, *74*, 1273–1298. https://doi.org/10.1007/s10640-019-00367-2

Altmann, A., Tolosi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347.

Basso, B., Ritchie, J. T., Cammarano, D., & Sartori, L. (2011). A strategic and tactical management approach to select optimal N fertilizer rates for wheat in a spatially variable field. *European Journal of Agronomy*, *35*(4), 215–222. https://doi.org/10.1016/j.eja.2011.06.004

Basso, B., Ritchie, J. T., Pierce, F. J., Braga, R. P., & Jones, J. W. (2001). Spatial validation of crop models for precision agriculture. *Agricultural Systems*, *68*(2), 97–112. https://doi.org/10.1016/S0308-521X(00)00063-9

Bebie, M., Cavalaris, C., & Kyparissis, A. (2022). *Assessing durum wheat yield through sentinel-2 imagery: A machine learning approach*. https://doi.org/10.3390/rs14163880

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, *151*, 61–69. https://doi.org/10.1016/j.compag.2018.05.012

Chollet, F., & Allaire, J. J. (2018). *Deep learning with R*. Manning Publications.

Colecchia, S. A., De Vita, P., & Rinaldi, M. (2015). Effects of tillage systems in durum wheat under rainfed Mediterranean conditions. *Cereal Research Communications*, *43*(4), 704–716. https://doi.org/10.1556/0806.43.2015.015

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.

Dasarathy, B. V. (1990). *Nearest neighbor (NN) norms: NN pattern classification techniques*. IEEE Computer Society Press.

Della Chiesa, S., Genova, G., & Niedrist, D. L. C. G. (2019). Phytoavailable phosphorus (P_2O_5) and potassium (K_2O) in topsoil for apple orchards and vineyards, South Tyrol, Italy. *Journal of Maps*, *15*(2), 555–562. https://doi.org/10.1080/17445647.2019.1633962

Denora, M., Fiorentini, M., Zenobi, S., Deligios, P. A., Orsini, R., Ledda, L., & Perniola, M. (2022). Validation of rapid and low-cost approach for the delineation of zone management based on machine learning algorithms. *Agronomy*, *12*(1). https://doi.org/10.3390/agronomy12010183

Diotallevi, F., Blasi, E., & Franco, S. (2015). Greening as compensation to production of environmental public goods: How do common rules have an influence at local level? The case of durum wheat in Italy. *Agricultural and Food Economics*, *3*(1), 17. https://doi.org/10.1186/s40100-015-0036-3

Feng, A., Zhou, J., Vories, E. D., Sudduth, K. A., & Zhang, M. (2020). Yield estimation in cotton using UAV-based multi-sensor imagery. *Biosystems Engineering*, *193*, 101–114. https://doi.org/10.1016/J.BIOSYSTEMSENG.2020.02.014

Filippi, P., Jones, E. J., Wimalathunge, N. S., Somarathna, P. D. S. N., Pozza, L. E., Ugbaje, S. U., Jephcott, T. G., Paterson, S. E., Whelan, B. M., & Bishop, T. F. A. (2019). An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. *Precision Agriculture*, *20*(5), 1015–1029. https://doi.org/10.1007/s11119-018-09628-4

Filippi, P., Whelan, B. M., Vervoort, R. W., & Bishop, T. F. A. (2020). Mid-season empirical cotton yield forecasts at fine resolutions using large yield mapping datasets and diverse spatial covariates. *Agricultural Systems*, *184*, 102894. https://doi.org/10.1016/J.AGSY.2020.102894

Fiorentini, M., Zenobi, S., Giorgini, E., Basili, D., Conti, C., Pro, C., Monaci, E., & Orsini, R. (2019). Nitrogen and chlorophyll status determination in durum wheat as influenced by fertilization and soil management: Preliminary results. *PLoS ONE*, *14*(11), e0225126. https://doi.org/10.1371/journal.pone.0225126

Fiorentini, M., Zenobi, S., & Orsini, R. (2021). Remote and proximal sensing applications for durum wheat nutritional status detection in Mediterranean area. *Agriculture*, *11*(1), 39. https://doi.org/10.3390/agriculture11010039

Friedman, J. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Gee, G. W., & Bauder, J. W. (1986). Particle-size analysis. In *Methods of soil analysis* (pp. 383–411). ASA, CSSA, and SSSA. https://doi.org/10.2136/sssabookser5.1.2ed.c15

Gómez, D., Salvador, P., Sanz, J., & Casanova, J. L. (2021). Modelling wheat yield with antecedent information, satellite and climate data using machine learning methods in Mexico. *Agricultural and Forest Meteorology*, *300*, 108317. https://doi.org/10.1016/J.AGRFORMET.2020.108317

Haghverdi, A., Washington-Allen, R. A., & Leib, B. G. (2018). Prediction of cotton lint yield from phenology of crop indices using artificial neural networks. *Computers and Electronics in Agriculture*, *152*, 186–197. https://doi.org/10.1016/J.COMPAG.2018.07.021

Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., & Zhang, J. (2020). Prediction of winter wheat yield based on multi-source data and machine learning in China. *Remote Sensing*, *12*(2), 236. https://doi.org/10.3390/rs12020236

Hijmans, R. J., van Etten, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., Lamigueiro, O. P., Bevan, A., Racine, E. B., & Shortridge, A. (2011). *Raster: Geographic data analysis and modeling* (R package version 2–0). https://cran.r-project.org/web/packages/raster/index.html

Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P. M., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., ... Keating, B. A. (2014). APSIM–Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, *62*, 327–350. https://doi.org/10.1016/J.ENVSOFT.2014.07.009

Hoogenboom, G., Jones, J. W., Wilkens, P. W., Porter, C. H., Batchelor, W. D., Hunt, L. A., Boote, K. J., Singh, U., Uryasev, O., & Bowen, W. T., Gijsman, A., Du Toit, A., White, J. W., Tsuji, G. Y. (2004). *Decision support system for agrotechnology transfer version 4.0 (CD-ROM)*. University of Hawaii.

Huang, Y., Brand, H. J., Sui, R., Thomson, S. J., Furukawa, T., & Ebelhar, M. W. (2016). Cotton yield estimation using very high-resolution digital images acquired with a low-cost small unmanned aerial vehicle. *Transactions of the ASABE*, *59*(6), 1563–1574. https://doi.org/10.13031/trans.59.11831

International Grains Council (IGC). (2020). *World grain statistics 2016*. IGC.

Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, *18*(3–4), 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7

Kayad, A., Sozzi, M., Gatto, S., Marinello, F., & Pirotti, F. (2019). Monitoring within-field variability of corn yield using sentinel-2 and machine learning techniques. *Remote Sensing*, *11*(23), 2873. https://doi.org/10.3390/rs11232873

Kheir, A. M. S., Alrajhi, A. A., Ghoneim, A. M., Ali, E. F., Magrashi, A., Zoghdan, M. G., Abdelkhalik, S. A. M., Fahmy, A. E., & Elnashar, A. (2021). Modeling deficit irrigation-based evapotranspiration optimizes wheat yield and water productivity in arid regions. *Agricultural Water Management*, *256*, 107122. https://doi.org/10.1016/J.AGWAT.2021.107122

Kim, N., & Lee, Y.-W. (2016). Machine learning approaches to corn yield estimation using satellite images and climate data: A case of

Iowa state. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, *34*(4), 383–390.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(1), 1–26.

Leo, S., Migliorati, M. D. A., & Grace, P. R. (2021). Predicting within-field cotton yields using publicly available datasets and machine learning. *Agronomy Journal*, *113*(2), 1150–1163. https://doi.org/10.1002/AGJ2.20543

Li, Q. Y., Yin, J., Liu, W. D., Zhou, S. M., Li, L., Niu, J. S., Niu, H. b., & Ma, Y. (2012). Determination of optimum growing degree-days (GDD) range before winter for wheat cultivars with different growth characteristics in North China Plain. *Journal of Integrative Agriculture*, *11*(3), 405–415. https://doi.org/10.1016/S2095-3119(12)60025-2

Li, Z., Cui, S., Zhang, Q., Xu, G., Feng, Q., Chen, C., & Li, Y. (2022). Optimizing wheat yield, water, and nitrogen use efficiency with water and nitrogen inputs in China: A synthesis and life cycle assessment. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/FPLS.2022.930484

Liang, L., Di, L., Zhang, L., Deng, M., Qin, Z., Zhao, S., & Lin, H. (2015). Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method. *Remote Sensing of Environment*, *165*, 123–134. https://doi.org/10.1016/J.RSE.2015.04.032

Lugato, E., Panagos, P., Bampa, F., Jones, A., & Montanarella, L. (2014). A new baseline of organic carbon stock in European agricultural soils using a modelling approach. *Global Change Biology*, *20*(1), 313–326. https://doi.org/10.1111/gcb.12292

Maestrini, B., & Basso, B. (2018). Drivers of within-field spatial and temporal variability of crop yield across the US Midwest. *Scientific Reports*, *8*(1), 14833. https://doi.org/10.1038/s41598-018-32779-3

Menegat, S., Ledo, A., & Tirado, R. (2022). Greenhouse gas emissions from global production and use of nitrogen synthetic fertilisers in agriculture. *Scientific Reports*, *12*, 14490. https://doi.org/10.1038/s41598-022-18773-w

Meng, L., Liu, H., Zhang, X., Ren, C., Ustin, S., Qiu, Z., Xu, M., & Guo, D. (2019). Assessment of the effectiveness of spatiotemporal fusion of multi-source satellite images for cotton yield estimation. *Computers and Electronics in Agriculture*, *162*, 44–52. https://doi.org/10.1016/J.COMPAG.2019.04.001

Mereu, V., Gallo, A., Trabucco, A., Carboni, G., & Spano, D. (2021). Modeling high-resolution climate change impacts on wheat and maize in Italy. *Climate Risk Management*, *33*, 100339. https://doi.org/10.1016/J.CRM.2021.100339

Mitchell, T. M. (1997). Artificial neural networks. *Machine Learning*, *45*, 81–127.

Nayak, H. S., Silva, J. V., Parihar, C. M., Krupnik, T. J., Sena, D. R., Kakraliya, S. K., Jat, H. S., Sidhu, H. S., Sharma, P. C., Jat, M. L., & Sapkota, T. B. (2022). Interpretable machine learning methods to explain on-farm yield variability of high productivity wheat in Northwest India. *Field Crops Research*, *287*, 108640. https://doi.org/10.1016/J.FCR.2022.108640

Nguyen, L. H., Zhu, J., Lin, Z., Du, H., Yang, Z., Guo, W., & Jin, F. (2019). Spatial-temporal multi-task learning for within-field cotton yield prediction. *Lecture Notes in Computer Science*, *11439*, 343–354. https://doi.org/10.1007/978-3-030-16148-4_27

Olsen, S. R. (1954). *Estimation of available phosphorus in soils by extraction with sodium bicarbonate (Issue 939)*. USDA.

Orsini, R., Basili, D., Belletti, M., Bentivoglio, D., Bozzi, C. A., Chiappini, S., Conti, C., Galli, A., Giorgini, E., Fiorentini, M.,

Malinverni, E. S., Mancini, A., Mazzanti, L., Monaci, E., Passerini, G., Pro, C., Santilocchi, R., Vignini, A., Zenobi, S., & Zingaretti, P. (2019). Setting of a precision farming robotic laboratory for cropping system sustainability and food safety and security: Preliminary results. *IOP Conference Series: Earth and Environmental Science*, *275*(1), 012021. https://doi.org/10.1088/1755-1315/275/1/012021

Orsini, R., Fiorentini, M., & Zenobi, S. (2020). Evaluation of soil management effect on crop productivity and vegetation indices accuracy in Mediterranean cereal-based cropping systems. *Sensors*, *20*(12), 3383. https://doi.org/10.3390/s20123383

Padilla, F. M., Farneselli, M., Gianquinto, G., Tei, F., & Thompson, R. B. (2020). Monitoring nitrogen status of vegetable crops and soils for optimal nitrogen management. *Agricultural Water Management*, *241*, 106356. https://doi.org/10.1016/j.agwat.2020.106356

Paudel, D., Boogaard, H., de Wit, A., van der Velde, M., Claverie, M., Nisini, L., Janssen, S., Osinga, S., & Athanasiadis, I. N. (2022). Machine learning for regional crop yield forecasting in Europe. *Field Crops Research*, *276*, 108377. https://doi.org/10.1016/j.fcr.2021.108377

Pro, C., Basili, D., Notarstefano, V., Belloni, A., Fiorentini, M., Zenobi, S., Alia, S., Vignini, A., Orsini, R., & Giorgini, E. (2021). A spectroscopic approach to evaluate the effects of different soil tillage methods and nitrogen fertilization levels on the biochemical composition of durum wheat (*Triticum turgidum* subsp. *durum*) leaves and caryopses. *Agriculture*, *11*(4). https://doi.org/10.3390/agriculture11040321

R. Core Team. (2014). *R: A language and environment for statistical computing*.

Rath, S., Zamora-Re, M., Graham, W., Dukes, M., & Kaplan, D. (2021). Quantifying nitrate leaching to groundwater from a corn-peanut rotation under a variety of irrigation and nutrient management practices in the Suwannee River Basin, Florida. *Agricultural Water Management*, *246*, 106634. https://doi.org/10.1016/j.agwat.2020.106634

Rhoades, J. D. (1983). Cation exchange capacity. In A. L. Page (Ed.), *Methods of soil analysis* (pp. 149–157). ASA, CSSA, and SSSA. https://doi.org/10.2134/agronmonogr9.2.2ed

Richetti, J., Judge, J., Boote, K. J., Johann, J. A., Uribe-Opazo, M. A., Becker, W. R., Paludo, A., & Silva, L. C., & de, A. (2018). Using phenology-based enhanced vegetation index and machine learning for soybean yield estimation in Paraná State, Brazil. *Journal of Applied Remote Sensing*, *12*(2), 1–15. https://doi.org/10.1117/1.JRS.12.026029

Rütting, T., Aronsson, H., & Delin, S. (2018). Efficient use of nitrogen in agriculture. *Nutrient Cycling in Agroecosystems*, *110*(1), 1–5. https://doi.org/10.1007/s10705-017-9900-8

Sagris, V., Wojda, P., Milenov, P., & Devos, W. (2013). The harmonised data model for assessing land parcel identification systems compliance with requirements of direct aid and agri-environmental schemes of the CAP. *Journal of Environmental Management*, *118*, 40–48. https://doi.org/10.1016/J.JENVMAN.2012.12.019

Schillaci, C., Perego, A., Valkama, E., Märker, M., Saia, S., Veronesi, F., Lipani, A., Lombardo, L., Tadiello, T., Gamper, H. A., Tedone, L., Moss, C., Pareja-Serrano, E., Amato, G., Kühl, K., Dămătîrcă, C., Cogato, A., Mzid, N., Eeswaran, R., . . . Acutis, M. (2021). New pedotransfer approaches to predict soil bulk density using WoSIS soil data and environmental covariates in Mediterranean agro-ecosystems. *Science of the Total Environment*, *780*, 146609. https://doi.org/10.1016/j.scitotenv.2021.146609

Seddaiu, G., Iocola, I., Farina, R., Orsini, R., Iezzi, G., & Roggero, P. P. (2016). Long term effects of tillage practices and N fertilization in rainfed Mediterranean cropping systems: Durum wheat, sunflower and maize grain yield. *European Journal of Agronomy*, *77*, 166–178. https://doi.org/10.1016/j.eja.2016.02.008

Shahhosseini, M., Hu, G., & Archontoulis, S. V. (2020). Forecasting corn yield with machine learning ensembles. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/FPLS.2020.01120

Shendryk, Y., Davy, R., & Thorburn, P. (2021). Integrating satellite imagery and environmental data to predict field-level cane and sugar yields in Australia using machine learning. *Field Crops Research*, *260*, 107984. https://doi.org/10.1016/j.fcr.2020.107984

Sparks, A. H. (2018). Nasapower: A NASA POWER global meteorology, surface solar energy and climatology data client for R. *The Journal of Open Source Software*, *3*(30), 1035. https://doi.org/10.21105/joss.01035

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*, *43*(6), 1947–1958. https://doi.org/10.1021/ci034160g

Valkama, E., Kunypiyaeva, G., Zhapayev, R., Karabayev, M., Zhusupbekov, E., Perego, A., Schillaci, C., Sacco, D., Moretti, B., Grignani, C., & Acutis, M. (2020). Can conservation agriculture increase soil carbon sequestration? A modelling approach. *Geoderma*, *369*, 114298. https://doi.org/10.1016/j.geoderma.2020.114298

van der Velde, M., & Nisini, L. (2019). Performance of the MARS-crop yield forecasting system for the European Usnion: Assessing accuracy, in-season, and year-to-year improvements from 1993 to 2015. *Agricultural Systems*, *168*, 203–212. https://doi.org/10.1016/J.AGSY.2018.06.009

Verhoeven, G. (2011). Taking computer vision aloft—Archaeological three-dimensional reconstructions from aerial photographs with photoscan. *Archaeological Prospection*, *18*(1), 67–73. https://doi.org/10.1002/arp.399

Walkley, A., & Black, I. A. (1934). An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science*, *37*(1), 29–38. https://doi.org/10.1097/00010694-193401000-00003

Wickham, H. (2016). *R eadxl: Read excel files*.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wu, L., Zhu, X., Lawes, R., Dunkerley, D., & Zhang, H. (2019). Comparison of machine learning algorithms for classification of LiDAR points for characterization of canola canopy structure. *International Journal of Remote Sensing*, *40*(15), 5973–5991. https://doi.org/10.1080/01431161.2019.1584929

Yang, J., Gong, W., Shi, S., Du, L., Sun, J., & Song, S. (2016). Laser-induced fluorescence characteristics of vegetation by a new excitation wavelength. *Spectroscopy Letters*, *49*(4), 263–267. https://doi.org/10.1080/00387010.2016.1138311

Zadoks, J. C., Chang, T. T., & Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Research*, *14*(6), 415–421. https://doi.org/10.1111/j.1365-3180.1974.tb01084.x

Zhang, C., & Ma, Y. (2012). *Ensemble machine learning: Methods and applications*. Springer Publishing Company.

Zhang, L., Zhang, J., Kyei-Boahen, S., & Zhang, M. (2010). Simulation and prediction of soybean growth and development under field conditions. *American-Eurasian Journal of Agricultural & Environmental Sciences*, *7*, 374–385.