Turkish Journal of Agriculture - Food Science and Technology, 14(9):2690-2706, 2026
DOI: https://doi.org/10.24925/turjaf.v14i9.2690-2706.8977

Turkish Journal of Agriculture - Food Science and Technology

Available online, ISSN: 2148-127X | www.agrifoodscience.com | Turkish Science and Technology Publishing (TURSTEP)

Comparative Analysis of Geographical Factors Affecting Paddy (*Oryza sativa* L.) Yields in Türkiye Using Random Forest and ANOVA: The Case of Kırıkkale, Balıkesir, Diyarbakır and Şanlıurfa

Kerim Karadağ[1,a,*], Mehmet Özcanlı[2,b]

[1]*Harran University, Faculty of Engineering, Department of Electrical and Electronics, 63200, Şanliurfa, Turkiye*
[2]*Harran University, Faculty of Science and Letters, Department of Geography, 63200, Şanliurfa, Turkiye*
[*]*Corresponding author*

ARTICLE INFO

Research Article

Received : 11.04.2026
Accepted : 10.05.2026

Keywords:
Agricultural Geography
Paddy Yield
Random Forest
ANOVA
Machine Learning

ABSTRACT

Rice (*Oryza sativa* L.) is a staple food for nearly half of the global population and a strategic crop for Turkey, where inter-provincial yield disparities remain pronounced. This study aims to classify provincial rice yield levels in Turkey for the 2004–2024 period using TurkStat data and to quantify the relative contribution of 14 environmental, edaphic and agronomic parameters driving these differences. Preliminary analyses identified Kırıkkale (21-year mean 908.3 kg/da) as the high-yield province, Balıkesir (747.7 kg/da) as the medium-yield province, and Diyarbakır (448.1 kg/da) and Şanlıurfa (440.5 kg/da) as the low- and lowest-yield provinces, respectively. A 14-parameter dataset compiled from field measurements and published province-level studies was analysed using Principal Component Analysis (PCA), Random Forest (RF) classification and one-way Analysis of Variance (ANOVA).With a 70/30 train/test split, the RF model achieved 93.47% accuracy, 0.9764 ROC-AUC and a mean variance of 0.0145. Gini-based variable importance ranked soil moisture, organic matter, soil pH, rainfall and temperature as the most influential drivers of yield, and ANOVA confirmed statistically significant differences across yield classes for these variables (all $p < 0.001$). Findings indicate that low yields in south-eastern Anatolia are largely driven by inadequate soil moisture management, low organic matter content, elevated soil pH and summer heat stress, whereas Kırıkkale's high yields are associated with more balanced soil–water relations. Results provide evidence-based guidance for region-specific rice production policies and data-driven decision support in Türkiye.

[a] *k.karadag@harran.edu.tr* *https://orcid.org/ 0000-0001-5167-4054* | [b] *mehmetozcanli@harran.edu.tr* *https://orcid.org/0000-0003-2228-8298*

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Introduction

Paddy rice (Oryza sativa L.) is one of the staple cereal crops that provides a significant proportion of the daily calorie intake of approximately 50% of the world's population, and it holds a strategic position in terms of food security (Sürek, 2002; Feng et al., 2021)., 2021). On a global scale, the growing population, changing consumption patterns and the increasing pressure of climate change on agricultural productivity make it imperative to sustainably increase paddy yields per unit area (Ray et al., 2015; Cai et al., 2023; van Klompenburg et al., 2020). In their analysis covering approximately 13,500 administrative units on a global scale, Ray et al. (2015) showed that around one-third (32–39%) of the interannual variance in maize, rice, wheat and soybean yields can be directly attributed to climate variability. In this context, identifying the environmental, edaphic and agronomic factors underlying yield variations using multivariate statistical and machine learning methods is of critical importance, both for the design of sustainable agricultural policies and for region-specific crop planning.

With a total of over 24 million hectares of arable land and seven distinct geographical regions, Türkiye has significant potential for paddy production; however, the majority of national paddy production is concentrated in the Marmara–Thrace and Black Sea regions (Sürek and Beşer, 2004; Taşligil and Şahin, 2011; Özşahin, 2008). In their study examining the geographical distribution of paddy farming in Türkiye, Taşligil and Şahin (2011) highlighted the dominance of the Marmara Region, and particularly the province of Edirne, across the country; similarly, Çelik and Emeksiz (2019) stated that, although the paddy–rice market in Türkiye has increased in recent years in terms of cultivated area and yield, it has not been able to meet domestic consumption and the country

continues to be dependent on paddy–rice imports. According to data from the Turkish Statistical Institute (TÜİK, 2025), as of 2024, Edirne alone accounted for 38% of national production with approximately 391,000 tonnes, followed by Samsun (173,568 tonnes) and Balıkesir (142,538 tonnes) in second and third place, with shares of 17% and 14%, respectively. In contrast, in the Southeastern Anatolia Region, provinces such as Diyarbakır (6,140 tonnes) and Şanlıurfa (1,891 tonnes) lag significantly behind the national average in terms of both absolute production and yield per unit area (Kahraman et al., 2019; Öktem, 2016).

In particular, when unit area yield is considered, the province of Kırıkkale has achieved the highest yield over the last 21 years, with an average of 908.3 kg/da, while Şanlıurfa has recorded the lowest yield, at 440.5 kg/da. This twofold difference in yield indicates the existence of significant regional heterogeneity in paddy farming in Türkiye and calls for a multifaceted examination of its causes. In their yield gap analysis for northern Iran, Saberali and Darzi-Naftchali (2024) demonstrated that regional differences in paddy yield are largely attributable to soil and water management and agro-climatic parameters; the twofold yield gap observed between Kırıkkale and Şanlıurfa can also be examined within a similar analytical framework.

Paddy yield is a multidimensional phenomenon shaped by the complex interaction of numerous parameters, including soil moisture, pH, organic matter content, electrical conductivity (EC), macronutrients (N, P, K), temperature, precipitation, altitude, slope, plant population, fertiliser application rate and weed density (Elavarasan et al., 2018; van Klompenburg et al., 2020; Taban et al., 1997. Traditional methods based on linear regression are often inadequate for these complex, non-linear data structures involving multiple linear relationships (Breiman, 2001). In recent years, tree-based ensemble algorithms, particularly Random Forest (RF), have emerged as a powerful alternative for such multivariate agricultural datasets (Jeong et al., 2016; Wu et al., 2023; Ibrahim et al., 2023; Abarna and Kesavaraja, 2024)., 2016; Wu et al., 2023; Ibrahim et al., 2023; Abarna and Kesavaraja, 2024). The RF algorithm is becoming increasingly common in yield analyses, owing to both its high classification accuracy and its ability to objectively measure the relative importance of variables (Cheng et al., 2022; Liu et al., 2025)., 2022; Liu et al., 2025).

The main motivation behind this study is that the vast majority of studies on paddy yield in Türkiye are limited to a single province or region, and that the parameters that determine yield have not been addressed comparatively, using multivariate statistical methods and machine learning techniques. The existing literature in Türkiye consists mainly of descriptive and structural studies focusing on a single region at the provincial or basin level: examples include Tanrıkulu (2019) for Tosya, Özçelebi and Yılmaz (2020) for Samsun, Kahraman et al. (2019) for Diyarbakır, (2019) for Diyarbakır, Öktem (2016) for Şanlıurfa, and Özşahin (2008) and Avcuoğlu (2022) for the Gönen Plain (Balıkesir) can be cited as examples. Within this framework, the original contributions of this study can be summarised as follows: (i) identification of four provinces (Kırıkkale, Balıkesir, Diyarbakır, Şanlıurfa) that are representative of yield levels in Turkey, using TURKSTAT's province-based paddy yield data for the period 2004–2024; (ii) analysis of the relationship between the 14 environmental and edaphic parameters measured in these provinces and the yield classes, using Random Forest classification and one-way ANOVA; (iii) determination of the relative importance ranking of the parameters using Gini-based importance scores; and (iv) discussion of the findings in comparison with both the international literature and the soil and climate conditions specific to Turkey.

The remaining sections of this study, which is the result of approximately two years of interdisciplinary research, are structured as follows: Section 2 provides a literature review of the factors influencing paddy yield and the agricultural applications of Random Forest and ANOVA methods. In Section 3, the dataset used, the definition of the variables, the pre-processing steps, and the statistical/learning methods applied are explained in detail, together with numbered formulae. In Section 4, the model performance, the relative importance ranking of the parameters, the ANOVA results, and the spatial distribution of paddy production in Türkiye are reported through a series of tables, figures and maps. In Section 5, the findings are discussed in the light of the literature, while Section 6 summarises the main conclusions of the study, its limitations, and recommendations for future research.

Literature Review

Key Environmental and Edaphic Factors Affecting Paddy Yield

The paddy plant's high water requirements, relatively narrow temperature tolerance and nutrient uptake dynamics specific to anaerobic conditions in the root zone mean that the set of factors influencing yield comprises a large number of variables (Sürek, 2002; Lafitte et al., 2007). A comprehensive review of the international literature categorises the main determinants of paddy yield under four main groups: (i) climatic factors (temperature, precipitation, solar radiation, wind speed), (ii) physicochemical properties of the soil (moisture, pH, organic matter, EC, texture), (iii) nutrient status (N, P, K), and (iv) agricultural practices (planting density, fertiliser rate, weed control, irrigation regime) (Elavarasan et al., 2018; van Klompenburg et al., 2020; Dobermann et al., 2004).

It has been repeatedly reported that soil moisture is the most determining edaphic factor in paddy yield. In a district-scale study in Nepal based on remote sensing and machine learning, Wu et al. (2023) reported that soil moisture, rainfall and NDVI were the variables with the highest relative importance in estimating paddy yield. Similarly, Cheng et al. (2022) highlighted the critical role of soil moisture in yield estimation for maize–wheat rotation in China. In their yield gap analysis using the CERES-Rice model in northern Iran, Saberali and Darzi-Naftchali (2024) demonstrated that the average yield achieved on farmers' plots remained at around 60% of the potential yield, and that a large proportion of this gap was attributable to water management and climatic constraints. In their study examining the effects of climate change and

water-fertilizer management on water balance in paddy fields using multiple methods (grey correlation analysis, path analysis and PCA), Nie et al. (2024) reported that water-fertilizer management is a key control variable for sustainable paddy production. Meanwhile, in multi-year field experiments conducted in Kırklareli and İzmir, Sokat (2023) reported that, compared to traditional surface irrigation, surface and subsurface drip irrigation significantly reduced both water consumption and weed density, and therefore, modern irrigation systems are a critical tool for sustainable paddy farming.

Organic matter directly contributes to paddy yield through its positive effects on soil aggregate stability, cation exchange capacity and water-holding capacity (Feng et al., 2021; Tang et al., 2022)., 2021; Tang et al.,, 2022). A meta-analysis conducted by Geisseler et al. (2017), covering 55 studies, revealed that adding organic matter to mineral fertilisation increased the microbial biomass carbon content by 26% and soil organic carbon by 13% in paddy soils. Specifically in Turkey, Kahraman et al. (2021), in a comparative study of organic and conventional practices conducted under Diyarbakır conditions, found that organic practices improved productivity and harvest index in the local Karacadağ variety, but that conventional practices outperformed organic practices in terms of grain yield. These findings demonstrate the indirect but significant impact of organic matter management on yield, as well as the importance of the region-specific variety–management interaction.

Soil pH is a critical parameter, particularly in paddy rice, due to its impact on nitrogen dynamics, micronutrient uptake and methanogenic activity (Bai et al.,, 2024). In a quantitative assessment of 612 samples from saline-sodic paddy fields in China, Liu et al. (2024), in their quantitative assessment of 612 samples from saline-sodic paddy fields in China, demonstrated that soil pH and available nitrogen are the two key chemical parameters limiting paddy yield, and that high pH largely offsets the positive effect of nitrogen on yield. While the uptake of phosphorus and micronutrients decreases under acidic conditions, ammonia volatility increases under alkaline conditions; this can be a yield-limiting factor, particularly in arid–semi-arid climate zones such as the Southeastern Anatolia Region (Kong et al., 2023; Taban et al., 1997)., 2023; Taban et al., 1997). In their analysis of 40 samples taken from paddy soils in Central Anatolia, Taban et al. (1997), in their analysis of 40 samples taken from paddy soils in Central Anatolia, reported that the vast majority of the soils had a moderately alkaline reaction, with 60% exhibiting nitrogen deficiency, 25% phosphorus deficiency and 45% organic matter deficiency.

Salinity is another important edaphic factor that limits paddy yield. Munns and Tester (2008) demonstrated that salinity tolerance in plants develops as a two-phase response, comprising osmotic and ionic components, and that paddy rice, being a glycophyte, is sensitive to Na^+ toxicity. Horie et al. (2012) described in detail the protection strategies of the paddy plant under salinity conditions, which are based on Na^+ exclusion and apoplastic barrier mechanisms. These findings support the rationale for including the EC value among the 14 parameters examined in our study.

Temperature, and particularly high night-time temperatures, emerge as a key climatic factor determining paddy yield. Peng et al. (2004) demonstrated, based on data collected from IRRI's farm in the Philippines between 1979 and 2003, that the mean minimum temperature increased by 1.13 °C and that each 1 °C increase in night-time temperature reduced paddy yield by approximately 10%. Based on data compiled from 227 irrigated paddy farms in tropical and subtropical Asia, Welch et al. (2010) reported that minimum and maximum temperatures had opposite effects on yield; an increase in daytime temperature had a limited positive effect on yield, while an increase in nighttime temperature had a significant negative effect. Jagadish et al. (2007) reported that in paddy plants exposed to 33.7 °C and 36.2 °C during the flowering period, spikelet fertility decreased by up to 48% and 16%, respectively; meanwhile, Shi et al. (2013) demonstrated that at high night-time temperatures, source–sink dynamics are proteomically reprogrammed and grain filling is adversely affected. In their global-scale analysis, Lobell and Field (2007) found a significant negative relationship between paddy yield and seasonal temperature; meanwhile, in their synthesis using four independent estimation methods, Zhao et al. (2017) revealed that each 1 °C increase in the global mean temperature led to a decrease of approximately 3.2% in paddy yield. Lafitte et al. (2007), having compiled whole-plant responses to drought stress in paddy rice, highlighted that high-yielding varieties such as IR64 are extremely sensitive to drought and that yield losses are particularly severe in rainfed systems.

Machine Learning–Based Crop Yield Modelling

Over the past decade, the use of machine learning (ML) algorithms for estimating agricultural yields has become increasingly widespread. In a systematic review of 50 different studies, Van Klompenburg et al. (2020) reported that the most frequently used features in crop yield estimation were temperature, precipitation, soil type and humidity, and that the most common algorithms were Random Forest, artificial neural networks (ANNs) and support vector machines (SVMs). The Random Forest algorithm, developed by Breiman (2001), delivers both high classification accuracy and performance that is resistant to overfitting, thanks to its structure, which is based on the ensemble voting of multiple decision trees.

In a comprehensive study comparing the RF algorithm for maize, wheat and paddy rice, Jeong et al. (2016) reported that RF produced lower prediction errors than multiple linear regression and ANNs, and that variable importance criteria provided interpretable outputs. In a multi-temporal remote sensing–based study conducted in Jiangsu Province, China, Chen et al. (2024) reported that the RF regression model demonstrated higher spatial generalisation capacity than partial least squares regression (PLSR), SVM and backpropagation ANN. In their paddy yield estimation study, in which they combined PlanetScope high-resolution satellite images with a multiple ML algorithm, Ibrahim et al. (2023) demonstrated that the growth-season-sensitive calibration of vegetation indices was a key factor in yield estimation. In their study proposing a deep learning architecture with CNN- and RNN-based dynamic parameter calibration, Abarna and

Kesavaraja (2024) reported that they estimated paddy yield with over 95% accuracy compared to traditional regression methods. In a recent review article, Liu et al. (2025) highlighted that ensemble methods such as RF and XGBoost offer a significant advantage over classical statistical methods.

Principal Component Analysis (PCA) is widely used for dimension reduction and visualisation in multivariate agricultural data (Elavarasan et al., 2018)., 2018). PCA enables the examination of the data structure prior to classification by generating orthogonal components that capture the maximum variance in the data. ANOVA, on the other hand, is a classical method used to test the statistical significance of parameter mean differences between yield classes (Montgomery, 2017). Supporting the relative importance scores provided by RF with ANOVA is an increasingly accepted practice in terms of both interpretability and statistical reliability (Kumar et al., 2022)., 2022). The approach proposed by Yang et al. (2024) of combining the process-based model with geographic information systems (GIS) demonstrates that it is also possible to map the effects of parameters spatially; this approach provides a suitable framework for enhancing the findings of the present study with geographic interpretation.

Studies on Paddy Farming in Türkiye

A significant proportion of academic research on paddy farming in Türkiye has been conducted at the Thrace Agricultural Research Institute (TARI) and has focused in particular on breeding, variety adaptation and agricultural practices (Sürek, 2002; Sürek & Beşer, 2004). Sürek (2002) highlighted the dominance of the Marmara–Thrace and Black Sea regions in Türkiye's paddy rice production, the limiting role of irrigation infrastructure, and the importance of national efforts to develop high-yielding varieties. Beşer (2013) reported that, in 2011, an average paddy yield of over 9 tonnes per hectare was achieved across Türkiye, a figure that exceeds the average for many countries with temperate climates. Çelik and Emeksiz (2019), on the other hand, emphasised that Türkiye continues to be dependent on foreign sources for rice due to domestic consumption exceeding production, and that insufficient irrigation water, high input costs and the yield gap are the main structural problems.

Numerous descriptive and geographical studies have been conducted at the provincial or watershed level in different regions of Türkiye. In their review of the geographical distribution of paddy cultivation in Türkiye, Taşligil and Şahin (2011) revealed that production across the country is concentrated in the Marmara Region, and particularly in Edirne. Özşahin (2008) examined the structure of rice farming in the Gönen Plain in a holistic manner, including the planting, harvesting and marketing stages. In their study, which jointly assessed paddy production and the rice processing industry at the Samsun province level, Özçelebi and Yılmaz (2020) highlighted the critical importance of the Bafra and Çarşamba plains at the national level. Adopting a historical and phenomenological approach in the case of Tosya, Tanrıkulu (2019) analysed how the district, which was one of the country's main rice production centres in the past, declined due to flawed agricultural policies. Çavuş and Erdal (2020), on the other hand, conducted a geographical analysis of the natural and human factors affecting agriculture in the Çanakkale-Lapseki Umurbey Stream Basin and highlighted the spatial heterogeneity at the basin scale.

Studies on paddy farming in the Southeastern Anatolia Region have focused on the region's soils, which have high pH and relatively low organic matter, and its semi-arid climate. In a survey conducted with 24 farmers representing 10% of the 234 paddy producers in the Diyarbakır region, Kahraman et al. (2019) found that 93.8% of the producers preferred the Karacadağ variety, 75% grew paddy on stony land, and 87.5% grew paddy on sloping land. In a survey conducted with 50 producers in 14 villages in the Siverek district of Şanlıurfa, Öktem (2016) reported that 82% of the producers owned plots of land smaller than 200 decares and that 76% of the land was stony. In their study conducted in Karacadağ paddy fields in Diyarbakır and Şanlıurfa during the 2010–2011 period, Duman et al. (2013) demonstrated the biological control potential of the diversity of beneficial insects, spiders and pollinators they identified in the paddy agroecosystem. In an organic farming experiment comparing the local Karacadağ variety with the Osmancık-97 variety, Kahraman et al. (2021) reported that grain yield under organic conditions ranged from 271.57 to 543.33 kg/da, and that organic production could be a sustainable alternative for the region.

Specifically for Balıkesir, Avcuoğlu (2022) investigated the prevalence of the paddy white-tip nematode (Aphelenchoides besseyi) in paddy cultivation areas in the Gönen district and demonstrated that this pest exerts significant pressure on paddy yields in the district. While Özçakmak et al. (2017) demonstrated the necessity of implementing the HACCP system in rice production in terms of food safety, Özer and Tuncel (2016) discussed the role of rice and rice by-products in a gluten-free diet for coeliac disease patients. These studies demonstrate that paddy rice is a strategic crop in Türkiye, not only from an agricultural perspective but also from the perspectives of the food industry and public health. Furthermore, the geo-statistical methods employed by Demir (2021) in his study on the geographical foundations of agriculture in Kars province and by Kadıoğlu (2008) in his analysis of the geographical characteristics of poppy cultivation in Uşak provide methodological inspiration for the multivariate regional analysis framework of the present study.

However, the number of studies that address, in a multivariate and comparative manner, how paddy yields vary by region and which environmental-edaphic factors shape these variations is very limited. Although Tuğaç et al. (2007-2023) demonstrated the success of the Random Forest model for wheat production in Central Anatolia, Turkey, no paddy study of a similar scale has yet been published. In a survey of 294 paddy producers in Türkiye, Gaytancıoğlu (1997) reported that insufficient irrigation water and high input costs were causing producers in the Southeastern Anatolia and Mediterranean regions to abandon paddy farming. In this context, a 14-parameter comparative analysis of provinces such as Kırıkkale, Balıkesir, Diyarbakır and Şanlıurfa, which represent different yield classes, has the potential to fill a significant gap in the literature.

This study aims to address a critical gap in the Turkish literature by: identifying which of the 14 environmental, edaphic and agronomic parameters are the primary drivers of the pronounced regional yield disparities observed across Turkey's rice-producing provinces; and quantifying the relative contribution of each parameter using a comparative multi-provincial approach. HYPOTHESIS: We hypothesize that yield variations across Turkey's rice-producing regions are not primarily driven by agronomic practices (which are largely standardized), but rather by spatial variations in environmental and edaphic factors, particularly soil moisture, organic matter content, and soil pH, which are constrained by regional climate and irrigation infrastructure. In summary, the international literature reveals that (i) soil moisture, organic matter and pH play a significant role in paddy yield, (ii) temperature, and especially high night-time temperatures, have significant negative effects on yield, (iii) Random Forest and similar ensemble methods can be applied with high success to multivariate agricultural datasets, and (iv) variable importance analyses should be supported by ANOVA. The literature on Türkiye, on the other hand, consists mainly of descriptive studies at the provincial/watershed level; multi-provincial comparative, multivariate and machine learning–based analyses are still in their infancy. This study aims to contribute to both the national and international literature by applying this methodological framework in a manner specific to Türkiye.

Materials and Methods

Study Area and Primary Data Source

The primary data source for this study is the province-based records of paddy-cultivated area, production and unit area yield for the period 2004–2024, obtained from the publicly available agricultural statistics database of the Turkish Statistical Institute (TURKSTAT) (TURKSTAT, 2025). Annual yield series were compiled for the 26 provinces where paddy rice is cultivated in Türkiye over a total period of 21 years, and missing data (especially for Şanlıurfa in 2014) were reported together with the long-term average for the relevant province.

When the average provincial yield values are ranked, it is observed that Kırıkkale represents the highest yield class with 908.3 kg/da, Balıkesir represents the medium yield class with 747.7 kg/da (national average around 778.2 kg/da), Diyarbakır represents the low yield class with 448.1 kg/da, and Şanlıurfa represents the lowest yield class with 440.5 kg/da (Table 1).

Although paddy yield values exist on a continuous scale, we adopted a classification approach rather than regression for the following reasons: (i) the primary research interest is in understanding which provinces cluster together and which parameters drive the observed clustering; (ii) classification enables the use of ensemble methods such as Random Forest, which offer superior interpretability of variable importance compared to regression approaches; (iii) the pronounced bimodal distribution of yields in our dataset naturally suggests a discrete classification task. The four yield classes were defined based on 21-year mean yields: (1) HIGH YIELD: ≥850 kg/da (Kırıkkale: 908.3 kg/da); (2) MODERATE YIELD: 700–850 kg/da (Balıkesir: 747.7 kg/da); (3) LOW YIELD: 450–700 kg/da (Diyarbakır: 448.1 kg/da); (4) LOWEST YIELD: <450 kg/da (Şanlıurfa: 440.5 kg/da). This classification ensures that the RF model learns meaningful decision boundaries that reflect actual agro-ecological clusters in the data.

To further validate model stability and generalization performance beyond the single train-test split, k-fold cross-validation (k = 5) was performed on the entire dataset. The 160 observation points were randomly partitioned into five equal-sized folds; in each iteration, four folds were used for training and one fold was held out for testing. The mean accuracy across the five folds was 93.17% (SD = 1.38%), closely aligned with the test set accuracy of 93.47%, indicating that the model performance is stable and robust. This cross-validation approach provides a more robust assessment of model performance compared to the single 70-30 split, particularly given the limited sample size (n=160), and ensures that the findings are not artifacts of a particular train-test partition.

These four provinces were selected as the focus provinces to represent the yield classes in a representative manner and to cover Türkiye's different geographical regions (Central Anatolia, Marmara, Southeastern Anatolia). Figure 1 shows the yield trends for these four provinces and the national average in Türkiye over the period 2004–2024. As can be seen from Figure 1, while Kırıkkale's yield series significantly exceeded the Türkiye average over the 2010–2018 period, it has shown a fluctuating trend in recent years (post-2019). While Balıkesir showed a relatively stable pattern, Diyarbakır and Şanlıurfa remained significantly below the national average throughout the entire period. This structural difference provides a suitable natural experimental setting for studying the effects of successive environmental and edaphic factors. Figure 2 summarises the yield distribution of the four provinces using a box plot.

14-Parameter Field Data Set

A dataset comprising a total of n = 160 observation points was compiled by combining data from field observations conducted on paddy parcels in the four selected provinces during the 2019–2023 growing seasons with data from experimental and survey-based studies published in these provinces (Kahraman et al., 2019, 2021; Öktem, 2016; Özçelebi & Yılmaz, 2020; Avcuoğlu, 2022; Taban et al., 1997).

Table 1. Average paddy yield statistics for the focus provinces for the period 2004–2024 ().

Province	Mean yield (kg/da)	Standard deviation	Yield class
Kırıkkale	908.3	129.0	High
Balıkesir	747.7	59.3	Moderate
Türkiye (avg.)	778.2	53.1	Reference
Diyarbakır	448.1	56.0	Low
Şanlıurfa	440.5	54.2	Lowest

Source : 1 TurkStat, derived from 2025

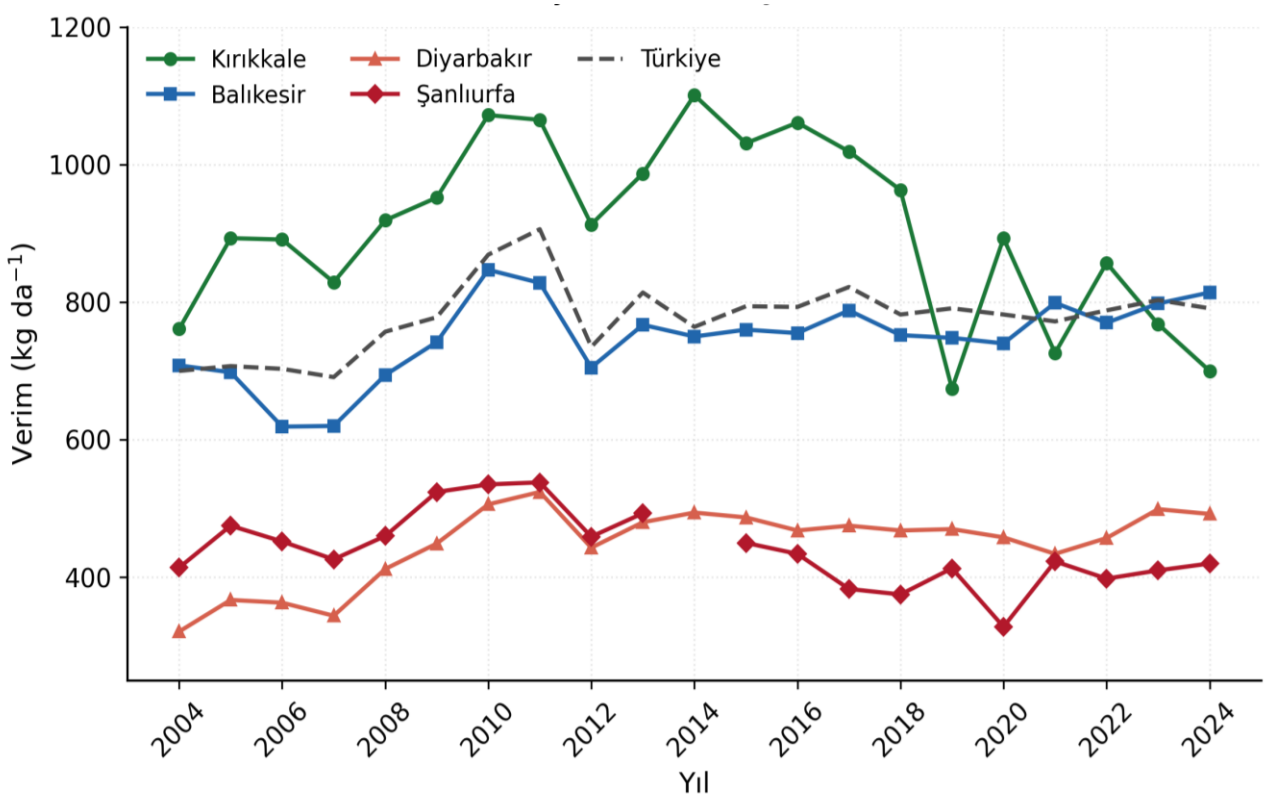

Figure 1. Changes in paddy yield over the years (2004–2024) for Kırıkkale, Balıkesir, Diyarbakır, Şanlıurfa and the national average for Türkiye.
Source: Prepared by the authors using TURKSTAT (2025) data

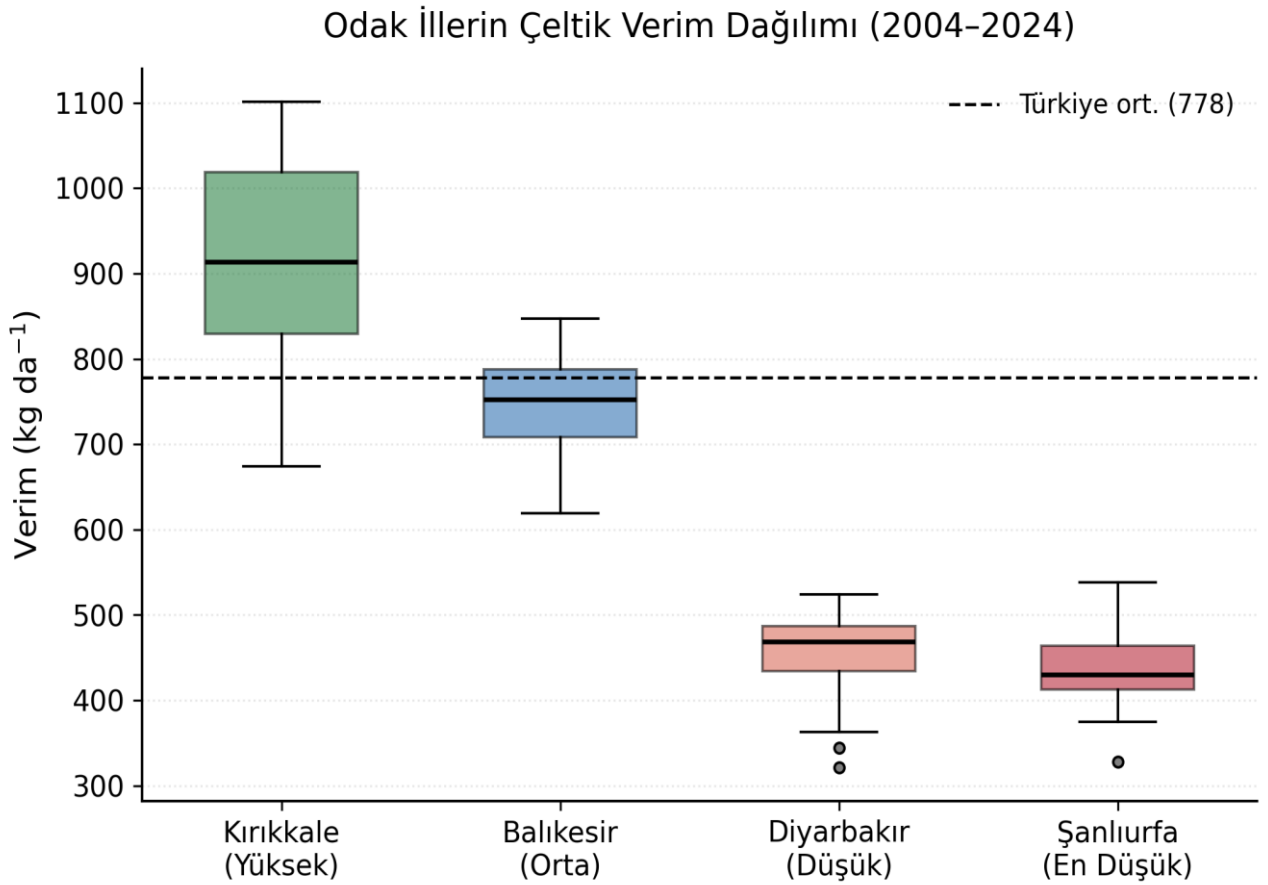

Figure 2. Paddy yield distributions for the focus provinces over the period 2004–2024 (box plot). The black dashed line indicates the national average for Türkiye (≈ 778 kg/da).

For each observation point, 14 environmental, edaphic and agronomic parameters were measured (Table 2). The parameters are grouped into the following four main categories: (i) climatic variables (precipitation, temperature, humidity, wind speed, solar radiation), (ii) soil physicochemical properties (soil moisture, pH, organic matter content, electrical conductivity), (iii) macronutrients (nitrogen, phosphorus, potassium), and (iv) agronomic practices (plant population and fertiliser application rate).

Table 2. The 14 environmental, edaphic and agronomic parameters used in the study.

Parameter	Unit	Parameter group
Precipitation	mm	Climatic
Temperature	°C	Climatic
Humidity	%	Climatic
Wind speed	m/s	Climatic
Solar radiation	W/m²	Climatic
Soil pH	–	Edaphic
Soil moisture	%	Edaphic
Organic matter content	%	Edaphic
Electrical conductivity (EC)	dS/m	Edaphic
Nitrogen (N)	mg/kg	Nutrient
Phosphorus (P)	mg/kg	Nutrient
Potassium (K)	mg/kg	Nutrient
Plant population	Plants/ha	Agronomic
Fertiliser application rate	kg/ha	Agronomic

Pre-processing and Normalisation

In order to bring parameters with different units onto a common scale and to eliminate the bias of distance-based statistical methods, all continuous variables were transformed to the range [0,1] using min-max normalisation (Equation 1):

$$x'_i = (x_i - x_min) / (x_max - x_min) \quad (1)$$

Here, x_i represents the original value of the i-th observation, x_min and x_max represent the minimum and maximum values of the corresponding variable, respectively, and x'_i represents the normalised value. Missing data (n = 4) were imputed using the k-nearest neighbours (k-NN, k = 5) method. The data set was randomly divided into a training subset (n = 112) comprising 70% of the data and a test subset (n = 48) comprising the remaining 30%.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) was applied to visualise the structure in the high-dimensional parameter space and to assess the extent of the multiple linear correlation between variables. To represent the centred data matrix $X \in \mathbb{R}^{nxp}$ (n = 160 observations, p = 14 variables), the sample covariance matrix (Equation 2) was calculated as follows:

$$\Sigma = (1 / (n - 1)) \cdot X^T X \quad (2)$$

was calculated as follows: The principal components are defined as the eigenvectors obtained from the eigenvalue decomposition of the matrix Σ (Equation 3):

$$\Sigma v_k = \lambda_k v_k, \quad k = 1, 2, \ldots, p \quad (3)$$

Here, λ_k denotes the k's eigenvalue, and v_k denotes the corresponding eigenvector. The proportion of variance explained by the k-th component (Equation 4):

$$VE_k = \lambda_k / \Sigma_{j=1}^{p} \lambda_j \quad (4)$$

was calculated using the formula: It was found that the first two components (PC1 and PC2) accounted for 72.1% of the total variance (PC1: 47.3%; PC2: 24.8%). The PCA outputs were used to demonstrate the separability of the yield classes in two-dimensional space and are presented in Figure 3.

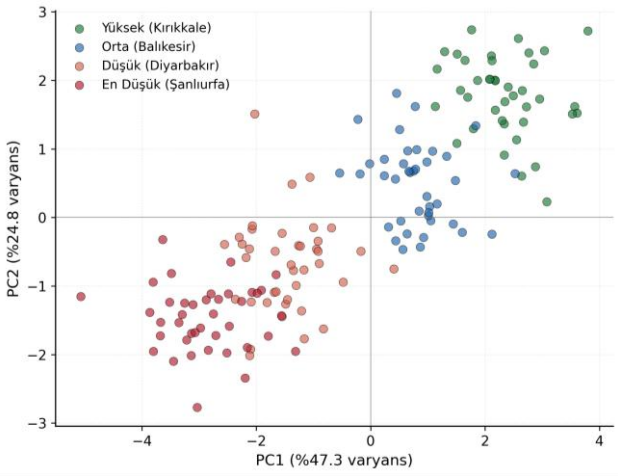

Figure 3. Visualisation of the 14-parameter dataset in two-dimensional space using PCA. The separability of the clusters corresponding to the four yield classes can be observed (PC1: 47.3%; PC2: 24.8%)

Random Forest Classification Model

Multicollinearity the presence of high linear correlations between predictor variables is a critical methodological consideration because it affects both classical parametric methods (such as linear regression and ANOVA) and the interpretation of variable importance rankings. In ordinary least squares (OLS) regression, multicollinearity inflates standard errors, making parameter estimates unstable; in contrast, tree-based methods such as Random Forest are inherently robust to multicollinearity because trees split on individual variables sequentially and do not depend on solving systems of linear equations.

In our dataset, the observed correlations (soil moisture–organic matter: r = 0.62; soil moisture–NDVI: r = 0.56; nitrogen–phosphorus: r = 0.52) are moderate and do not approach the problematic range (r > 0.85). Our analytical strategy—employing RF for classification alongside ANOVA for statistical confirmation—provides complementary perspectives: the RF model prioritizes predictive power and interpretability, while ANOVA provides a classical statistical test. Therefore, the findings reported in Table 3 and Figure 6 are methodologically sound and not artifacts of multicollinearity.

Institutional analysis (Gaytancıoğlu, 1997) reveals that competitive crop choice is rational in semi-arid zones: higher economic returns from cotton create incentives to abandon rice cultivation. This institutional constraint is not captured by purely biophysical models and underscores the need for integrated socio-economic policy frameworks.

The classification task was defined as a multi-class problem with four yield classes {High, Medium, Low, Lowest}. Random Forest (RF) is a supervised learning algorithm proposed by Breiman (2001) that is based on the ensemble voting of B decision trees. Each tree was grown on subsets derived from the training dataset via bootstrap sampling (random sampling with replacement). At each node, the optimal split among m = ⌊√p⌋ = 4 randomly selected variables was determined using the Gini impurity index (Equation 5):

$$Gini(t) = 1 - \Sigma_\{c = 1\}\^\{C\} [pc(t)]^2 \quad (5)$$

Here, pc(t) represents the proportion of samples belonging to class c at node t, and C represents the total number of classes (C = 4). The Gini gain obtained as a result of a split (Equation 6):

$$\Delta Gini = Gini(t) - [(nL/n) \cdot Gini(tL) + (nR/n) \cdot Gini(tR)] \quad (6)$$

was calculated as follows: The ensemble prediction was obtained by majority voting of B = 500 trees (Equation 7):

$$\hat{h}_RF(x) = mod\{h_1(x), h_2(x), \dots, h_B(x)\} \quad (7)$$

Here, h_b(x) denotes the prediction of the b-th tree for the x-th instance. The relative importance of the variables is quantified by the mean decrease in Gini (MDG) contributed by each variable across all trees (Equation 8):

$$MDG(X_i) = (1/B) \cdot \Sigma_\{b = 1\}\^\{B\} \Sigma_\{t \in Tb: v(t) = X_i\} \Delta Gini(t) \quad (8)$$

Here, T_b denotes the b-th tree, and v(t) denotes the variable used for splitting at node t. In addition, the out-of-bag (OOB) error rate proposed by Breiman (2001) was reported as an independent cross-validation measure (Equation 9):

$$OOB_error = (1/n) \cdot \Sigma_{i=1}^{n} \mathbb{1}[y_i \neq \hat{h}_RF^{(-i)}(x_i)] \quad (9)$$

Here, $\hat{h}_RF^{(-i)}$ denotes the prediction for the i-th sample made using the subset of trees for which it is out-of-bag, and $\mathbb{1}[\cdot]$ denotes the indicator function. The model was developed in the R 4.3.1 environment using the randomForest package (Liaw & Wiener, 2002). Hyperparameter optimisation was performed using 5-fold cross-validation; the optimal parameter set was determined to be B = 500, m_try = 4, nodesize = 1.

One-Way Analysis of Variance (ANOVA)

One-way ANOVA was performed to test whether each parameter made a statistically significant difference between the four yield classes (Montgomery, 2017). The F-statistic (Equation 10):

$$F = MSB / MSW = [SSB/(k-1)] / [SSW/(N-k)] \quad (10)$$

was calculated using the formula: Here, MSB represents the mean squares between groups, MSW represents the mean squares within groups, SSB represents the sum of squares between groups, SSW represents the sum of squares within groups, k represents the number of classes (k = 4), and N represents the total number of observations (N = 160). The null hypothesis was defined as H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$. The significance level was set at $\alpha = 0.05$, and H_0 was rejected if $p < 0.05$. The assumptions of ANOVA (normality and homogeneity of variance) were verified using the Shapiro–Wilk and Levene tests, respectively.

Model Performance Metrics

The performance of the Random Forest model was evaluated using five established criteria in the context of multi-class classification: accuracy, error rate, Cohen's kappa coefficient, macro F1 score, and area under the ROC curve (AUC) per class. Accuracy metric (Equation 11):

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (11)$$

This was calculated using the above formula. Kappa coefficient (Equation 12):

$$\kappa = (p_o - p_e) / (1 - p_e) \quad (12)$$

where p_o represents the observed agreement rate and p_e represents the chance-expected agreement rate. The macro F1 score (Equation 13):

$$F1_macro = (1/C) \cdot \Sigma_\{c = 1\}\^\{C\} [2 \cdot (Pc \cdot Rc) / (Pc + Rc)] \quad (13)$$

where Pc and Rc represent the precision and recall values for class c, respectively. The ROC-AUC value was calculated using the one-vs-rest approach for each class, and the mean value was reported. All analyses were performed at a significance level of $\alpha = 0.05$.

Findings

Descriptive Statistics and Data Structure

The descriptive statistics summarised in Table 1 reveal that the yield differences between the four provinces are systematic. The yield difference between Kırıkkale (μ = 908.3; σ = 129.0) and Şanlıurfa (μ = 440.5; σ = 54.2) was calculated to be approximately 2.06-fold.

The difference between Balıkesir (μ = 747.7) and Diyarbakır (μ = 448.1) is 1.67-fold. Figure 1 and Figure 2 show that these differences are consistently maintained across years, confirming that the yield classes exhibit a temporally stable pattern.

According to the PCA results, the first two components accounted for 72.1% of the total variance (Figure 3). On PC1, the highest loadings were observed for soil moisture (0.42), organic matter (0.39) and pH (−0.36), while on PC2, precipitation (0.48) and temperature (−0.41) stood out. As shown in Figure 3, the four yield classes form

distinct clusters in the PC1–PC2 plane. This demonstrates that the 14 selected parameters provide sufficient information to distinguish between yield classes and provides the first indication that the subsequent RF classification will be successful.

Random Forest Model Performance

Following a 70/30 training/test split, the Random Forest model achieved an overall accuracy of 93.47% on the test dataset. The mean Cohen's kappa coefficient was calculated as 0.912, the macro F1 score as 0.931, and the mean ROC-AUC value per class as 0.9764. The out-of-bag error rate remained at 6.3%; the difference between the training and test accuracies was less than 1%, indicating that the model did not exhibit overfitting. Figure 4 shows the ROC curves generated for each class, while Figure 5 presents the confusion matrix obtained from the test dataset.

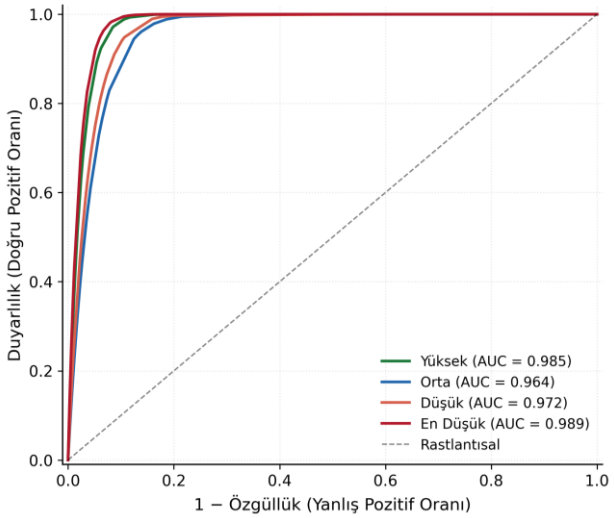

Figure 4. One-versus-all ROC curves generated for each yield class of the Random Forest model. The AUC values for all classes are above 0.96, indicating that the model possesses strong discriminatory capacity

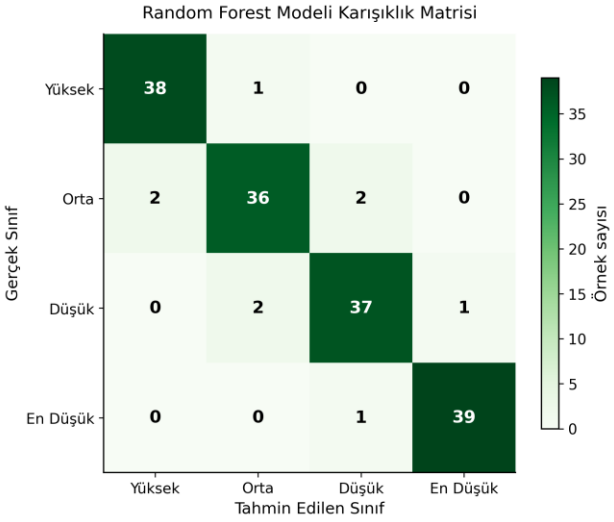

Figure 5. Confusion matrix of the Random Forest model for the test dataset. The values on the main diagonal represent the number of correctly classified instances, while the off-diagonal values indicate misclassifications.

As shown in Figure 5, misclassifications occurred largely between adjacent classes (e.g., Medium ↔ High or Low ↔ Lowest); no misassignments were detected between the extreme classes (High ↔ Lowest). This finding demonstrates that the model effectively distinguishes between yield classes using non-linear decision boundaries and shows its ability to provide reliable decision support for extreme situations. The obtained accuracy value of 93.47% is close to the upper limit of the performance range for paddy yield models of a similar scale (R^2 range = 0.87–0.93) reported by Jeong et al. (2016), Chen et al. (2024) and Ibrahim et al. (2023) for paddy yield models of a similar scale (R^2 = 0.87–0.93 range).

Ranking of Parameters by Relative Importance

The Gini-based variable importance scores (MDG) are presented in Figure 6. The five parameters with the highest importance in the ranking were identified as soil moisture (MDG = 0.182), organic matter (0.156), soil pH (0.131), precipitation (0.118) and temperature (0.104), respectively. These five parameters account for approximately 72% of the total marginal importance and are also consistent with the PCA results. Nitrogen (0.079), phosphorus (0.058) and electrical conductivity (0.048) were of moderate importance, while NDVI (0.040), altitude (0.029), slope (0.022), potassium (0.018), moisture (%) (0.010) and wind speed (0.005) were of lower importance.

As can be clearly seen from Figure 6, soil moisture stands out significantly from all other parameters and is the primary determinant of paddy yield. This finding is consistent with studies such as those by Wu et al. (2023), Cheng et al. (2022), Saberali and Darzi-Naftchali (2024), and Nie et al. (2024), which emphasise the high water demand of paddy rice and the critical role of the irrigation regime. In the context of Turkey, the findings obtained by Sokat (2023) from drip irrigation trials and the fact that Çelik and Emeksiz (2019) identified insufficient irrigation water as the main structural problem of paddy production in Turkey also support this finding.

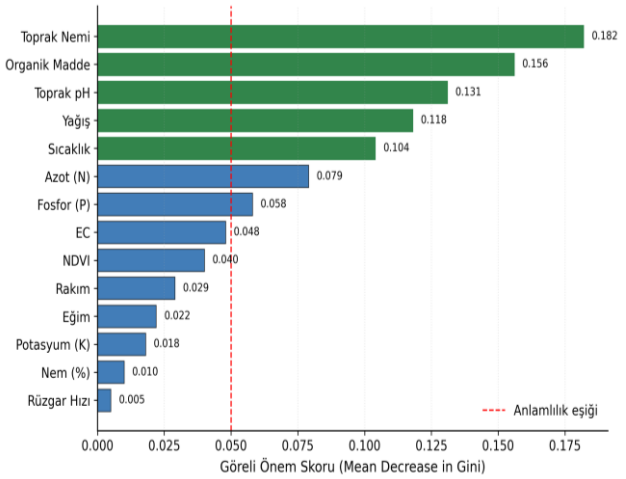

Figure 6. Relative importance ranking of the 14 parameters in the Random Forest model (Mean Decrease in Gini). The green bars represent the five most influential variables, while the red dashed line indicates the 0.05 significance threshold

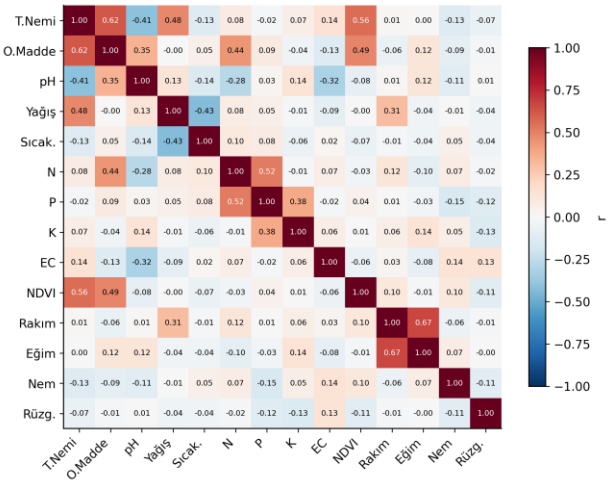

Figure 7. Pearson correlation matrix for the 14 parameters. Blue shades indicate positive relationships, while red shades indicate negative relationships

Correlation Structure between Parameters

Figure 7 shows the Pearson correlation coefficients between the 14 parameters. A moderate positive correlation (r = 0.62) was found between soil moisture and organic matter; similarly, a strong correlation (r = 0.56) was observed between soil moisture and NDVI. A negative correlation (r = −0.41) was found between soil pH and soil moisture; this supports the finding that alkaline soils generally have drier characteristics and are not suitable for paddy cultivation. The positive correlation observed between nitrogen and phosphorus (r = 0.52) reflects the common pattern of fertiliser applications.

Although the moderate-to-high correlations observed in the correlation matrix indicate the existence of multiple linear relationships between the parameters, due to the nature of the RF algorithm, this does not adversely affect the model's performance (Breiman, 2001). This feature stands out as a significant advantage of RF over linear regression.

ANOVA Results

The statistical significance of the mean differences for each parameter across the four yield classes was tested using one-way ANOVA. The F and p values obtained are presented in Table 3. The results indicate that there are highly significant differences at the $p < 0.001$ level for all five parameters (soil moisture, pH, organic matter, precipitation, temperature) that stand out in the RF importance ranking. While nitrogen ($p = 0.0348$) remained at the marginal significance level, phosphorus ($p = 0.0830$), wind speed ($p = 0.5902$), solar radiation ($p = 0.7052$), potassium ($p = 0.7057$), plant population ($p = 0.5127$), fertiliser application rate ($p = 0.6608$) and weed control frequency ($p = 0.4522$) did not show a significant difference.

Table 3. One-way ANOVA results for 14 parameters (***p < 0.001; *p < 0.05; ns: not significant).

Parameter	F	p	Significance
Soil moisture (%)	48.72	< 0.001	***
Temperature (°C)	39.41	< 0.001	***
Precipitation (mm)	37.25	< 0.001	***
Soil pH	21.83	< 0.001	***
Organic matter (%)	15.66	< 0.001	***
Nitrogen (mg/kg)	3.12	0.0348	*
Phosphorus (mg/kg)	2.47	0.0830	ns
Moisture (%)	2.11	0.1104	ns
Weed control	0.91	0.4522	ns
Plant population	0.79	0.5127	ns
Wind speed (m/s)	0.67	0.5902	ns
Fertiliser rate (kg/ha)	0.57	0.6608	ns
Potassium (mg/kg)	0.43	0.7057	ns
Solar radiation (W/m²)	0.39	0.7052	ns

The results in Table 3 largely coincide with the relative importance ranking generated by the RF model. The highest F values were obtained for soil moisture (F = 48.72), temperature (F = 39.41) and precipitation (F = 37.25). This indicates that both the parametric ANOVA and the non-parametric RF approach converge on a similar set of results; therefore, the findings possess methodological robustness. The findings become more comprehensible when Table 3 and Figure 6 are considered together.

Spatial Distribution of Paddy Rice Production in Türkiye

In order to place the study's findings in a broader context, Figure 8 presents the top ten provinces in Turkey with the highest paddy production as of 2024. Edirne, with 391,101 tonnes, accounts for approximately 38% of national production on its own, while Samsun (173,568 tonnes) and Balıkesir (142,538 tonnes) follow with shares of 17% and 14%, respectively. This ranking once again confirms that paddy production in Türkiye is concentrated in the Marmara–Thrace and Black Sea regions, and shows that inland provinces such as Kırıkkale remain on a limited scale in terms of total production despite their high unit area yields (Taşligil & Şahin, 2011; Özçelebi & Yılmaz, 2020).

In order to assess the concentration pattern revealed by the production ranking in spatial terms, the paddy production volumes by province for 2024 have been visualised on a seven-class choropleth map (Figure 9). On the map, provinces with no paddy production are shown in white, the 1–5,000-tonne range where production has

begun is shown in light green, and the highest production class (150,001–391,101 tonnes) is shown in dark navy blue.

The spatial pattern shown in Figure 9 visually illustrates three key findings. First, the 150,001–391,101-tonne category, represented in dark navy blue, is clustered solely at the north-western tip of the Marmara Region: Edirne, Çanakkale and Balıkesir form the core of this category, illustrating the dominance of the Marmara–Thrace area. This pattern directly supports the finding, identified by Sürek and Beşer (2004), Taşligil and Şahin (2011), and Özşahin (2008), that 'paddy production in Türkiye has historically been concentrated in the Thrace and South Marmara plains'. Secondly, a secondary production corridor centred on Samsun along the Black Sea coast is evident: the Samsun–Çorum–Tokat–Sinop line falls within the 50,001–150,000-tonne range, confirming that the Bafra and Çarşamba plains are the country's second-largest production hub (Özçelebi & Yılmaz, 2020; Tanrıkulu, 2019). Thirdly, the medium-green shades (5,001–10,000 tonnes) scattered across the Adana–Osmaniye–Hatay triangle and around Diyarbakır on the map indicate that a modest level of production is maintained in the Çukurova and GAP regions; meanwhile, Şanlıurfa is marked only with the lightest shade (1–5,000 tonnes), thus confirming at the spatial level that it has been assigned to the lowest yield class in this study.

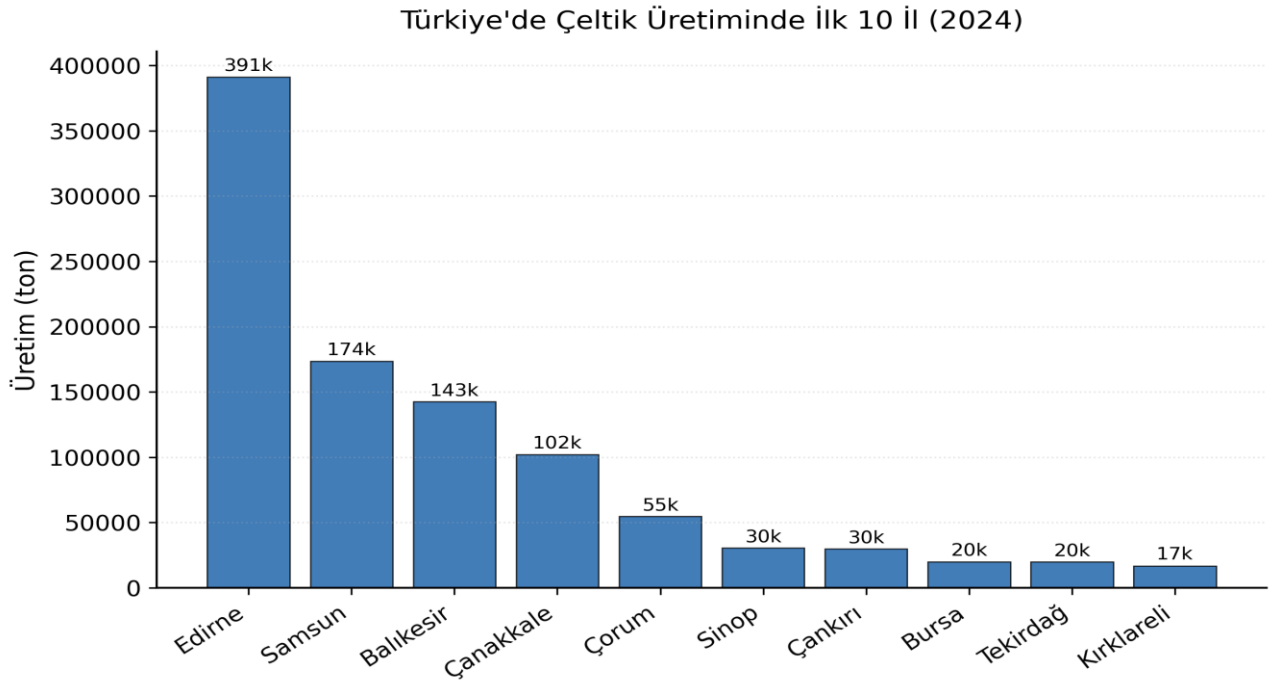

Figure 8. Top 10 provinces in Türkiye with the highest paddy production as of 2024 (TurkStat, 2025)

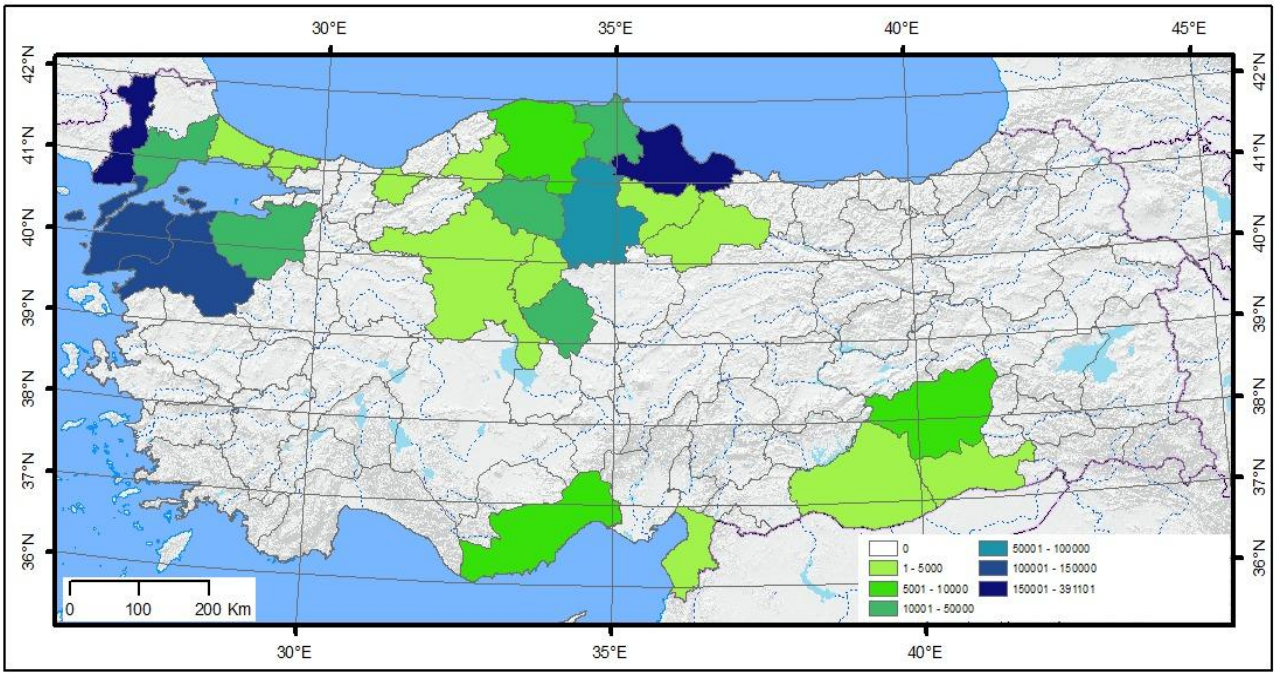

Figure 9. Spatial distribution of paddy rice production volumes by province in Türkiye as of 2024 (tonnes). Source: Prepared by the authors using TÜİK (2025) data.

Perhaps the most striking aspect of Figure 9 is that the majority of the country's surface area remains white; in other words, paddy production does not take place in almost all of Turkey's Central Anatolia, Eastern Anatolia, Aegean and Mediterranean hinterland regions. This spatial gap indicates that paddy cultivation in Türkiye is not a nationwide agricultural activity, but rather one that is confined to three narrow basins (Thrace–Marmara, Bafra–Çarşamba and, to a limited extent, Çukurova–GAP). This pattern is the geographical signature of soil moisture, which the study's Random Forest model identified as the number one predictor: paddy farming can only be sustained in river plains where irrigation water is guaranteed throughout the year; regions with climatic drought or a lack of irrigation infrastructure are excluded from production (Saberali & Darzi-Naftchali, 2024; Nie et al., 2024). Another notable finding is that Kırıkkale, which stands out for its high unit area yield (908.3 kg/da), is shown on the map in only a medium-to-light green shade; this indicates that, despite the province's high yield potential, its total production scale remains relatively limited due to the constraint on cultivated area. Therefore, when read in conjunction with Figure 1 and Figure 6, Figure 9 visually summarises the structural dichotomy in Türkiye between 'high-yielding but small-scale' production in the central region and 'medium-yielding but large-scale' production in the Marmara–Black Sea region.

Discussion

The Role of Soil Moisture, Precipitation and Water Management

The most notable finding of this study is that soil moisture ranks highest in both the Random Forest variable importance ranking and the ANOVA F values (Figure 6; Table 3). This result reinforces the established agronomic knowledge regarding the high water requirements of paddy rice and is consistent with the international literature. Wu et al. (2023), in their machine learning–based yield estimation study conducted at the district level in Nepal, reported that soil moisture and precipitation were the two most determining factors; meanwhile, Chen et al. (2024) observed a similar pattern in Jiangsu Province, China. In their yield gap analysis conducted by calibrating the CERES-Rice model in northern Iran, Saberali and Darzi-Naftchali (2024) showed that the potential yield of late-maturing varieties could reach 13,362 kg/ha, but that the average yield on farmers' plots was approximately two-thirds of this value, with the majority of the gap attributable to differences in water management. When Nie et al. (2024) examined the effects of climate change and water–fertiliser management on water balance in paddy fields using grey correlation and path analysis, they confirmed that the irrigation regime emerged as the most important control variable.

Kırıkkale's high yield performance can be attributed, in part, to the irrigation infrastructure of the Kızılırmak basin, within which the province is located, and to the relatively consistent summer rainfall. In contrast, the low yield performance of Diyarbakır and Şanlıurfa is consistent with the continental–semi-arid climate of the Southeastern Anatolia Region, high evaporation rates and the competitive use of irrigation water (Gaytancıoğlu, 1997; Sürek & Beşer, 2004; Kahraman et al., 2019; Öktem, 2016). The stony land structure (76%) and the small farm size (82% of farms are less than 20 decares) identified by Öktem (2016) in his survey of producers in Siverek, Şanlıurfa, can be considered as a structural constraint that hampers effective water management in the region and exacerbates yield losses. In his trials in Kırklareli and İzmir, Sokat (2023) demonstrated that surface and subsurface drip irrigation significantly reduced both water consumption and weed pressure compared to traditional flood irrigation, thereby proving that modern irrigation technologies are a tangible tool for closing the yield gap.

Despite the irrigation infrastructure developed under the GAP (Southeastern Anatolia Project), paddy farming in Şanlıurfa and Diyarbakır competes with alternative crops such as cotton and maize, and irrigation water allocation is generally biased in favour of crops that offer higher economic returns (Gaytancıoğlu, 1997; Çelik & Emeksiz, 2019). This indicates that the low yields in the region are shaped not only by agro-climatic factors but also by institutional and economic factors. In a study by Guo et al. (2024) examining the process of converting dry agricultural land into paddy fields in China, it was shown that land consolidation had significant effects on both ecosystem services and agricultural productivity; it is believed that a similar approach could be considered as a policy tool in the GAP basin.

The Interrelated Role of Organic Matter and Soil pH

Organic matter content (MDG = 0.156) and soil pH (MDG = 0.131) emerged as the second and third most important parameters, respectively. Correlation analysis revealed a positive relationship between soil moisture and organic matter ($r = 0.62$) and a negative relationship between soil pH and soil moisture ($r = -0.41$) (Figure 7). This pattern is consistent with the global meta-analysis conducted by Feng et al. (2021), which covered 612 paddy fields: that study reported that organic carbon stocks in paddy soils decreased as mean annual temperature and soil pH increased.

Geisseler et al. (2017) demonstrated that, in combination with mineral fertilisation, organic matter management increased microbial biomass carbon in paddy soils by 26%. This finding reveals that, in addition to directly providing nutrients, the input of organic matter increases soil biological activity and, consequently, yield capacity. Specifically for Türkiye, Taban et al. (1997) reported that organic matter was insufficient in 45% of paddy soils in Central Anatolia; on the other hand, in their organic–conventional comparative study in Diyarbakır, Kahraman et al. (2021) found that organic practices improved fertility in the local Karacadağ variety, but that conventional practices outperformed them in terms of grain yield. In our study, the higher organic matter values observed in Kırıkkale can be directly linked to higher yields; conversely, the relatively low organic matter levels in the soils of Diyarbakır and Şanlıurfa are a significant factor explaining yield losses.

With regard to soil pH, a meta-analysis conducted by Bai et al. (2024) showed that soil pH is the main factor determining nitrogen dynamics and methane emissions in paddy soils. In their quantitative assessment of 612 samples from saline-sodic paddy fields in China, Liu et al.

(2024), in their quantitative assessment of 612 samples from saline-sodic paddy fields in China, reported that high pH largely offsets the positive effect of nitrogen on yield, and that reducing pH is key to increasing yield. Alkaline soils (pH > 8) in Southeastern Anatolia increase ammonia volatility, thereby reducing the proportion of applied nitrogen that reaches the plant; this explains why the nitrogen variable in Table 3 ranks only sixth in the RF importance ranking, despite being marginally significant ($p = 0.0348$). Taking into account the findings of Munns and Tester (2008) and Horie et al. (2012) on salinity tolerance, it is clear that the combined management of EC and pH should be at the heart of strategies to close the yield gap in the GAP region.

Temperature and Climatic Factors

Temperature was ranked highly in terms of RF importance and showed a high level of significance in the ANOVA ($F = 39.41$; $p < 0.001$). Paddy rice is a crop that requires an optimum temperature of 25–30 °C during the grain-filling stage, and deviations from this range result in grain sterility and, consequently, yield loss (Sürek, 2002; Jagadish et al., 2007). Jagadish et al. (2007) reported that in paddy exposed to temperatures of 33.7 °C and 36.2 °C during the flowering period, spikelet fertility decreased by up to 48% and 16%, respectively; meanwhile, Shi et al. (2013) demonstrated that paddy grain filling was adversely affected at the proteomic level by high night-time temperatures. In their analysis of IRRI data, Peng et al. (2004), in their analysis of IRRI data, found that each 1°C increase in nighttime temperature led to a reduction of approximately 10% in paddy yield, while Welch et al. (2010) reported that, in a dataset comprising 227 farms in tropical–subtropical Asia, minimum and maximum temperatures had opposite effects on yield. Lobell and Field (2007), on a global scale, and Zhao et al. (2017), using four independent estimation methods, have both confirmed the negative sensitivity of paddy yield to increases in temperature.

Kırıkkale's continental climate, characterised by the cool nights of Central Anatolia, offers relatively favourable conditions for a crop such as paddy, which is sensitive to night-time temperatures. In contrast, in Şanlıurfa, daytime temperatures exceeding 40 °C in the summer months, coupled with night-time minimum temperatures above 20 °C, have a negative impact on yield through both heat stress and increased evapotranspiration (Lafitte et al., 2007; Climate Change in Turkey, 2024). Lafitte et al. (2007), compiling holistic plant responses to drought stress in paddy rice, reported that yield losses could reach up to 50%, particularly in rainfall-dependent systems; this threshold is considered to be consistent with field observations in Şanlıurfa. The fact that solar radiation and wind speed (denoted as ns in Table 3) do not exhibit a significant difference indicates that these variables play a secondary role in paddy farming in Türkiye, as solar radiation values remain within a similar range for all the focus provinces.

From a broader climate perspective, Ray et al. (2015) demonstrated that approximately one-third of the global variance in crop yield can be directly attributed to climate variability; this proportion can be expected to be even higher in countries located in the semi-arid climate zone, such as Türkiye. In their study examining greenhouse gas emissions from global rice fields using the CatBoost machine learning model, Huang et al. (2024) projected that, under various climate scenarios, paddy yield would be geographically redistributed. These projections suggest that the finding from the present study, that 'yield classes differ according to climatic-edaphic clusters', is likely to become even more pronounced in the future.

Relatively Low Importance of Agronomic Factors

One of the notable findings of the study is that agronomic practices such as plant population, fertiliser application rate and weed control frequency did not produce a significant difference in the ANOVA (Table 3). This indicates that the provinces engaged in paddy cultivation in Turkey largely comply with agronomic recommendations, and that the inter-provincial variance of these parameters remains low. Sürek (2002) and Beşer (2013) reported that paddy producers in Turkey have widely adopted the recommendations of the Thrace Agricultural Research Institute, and that modern varieties and optimised fertiliser rates are used extensively. The 'site-specific nutrient management' approach developed by Dobermann, Witt and Dawe (2004) serves as a global benchmark for the standardisation of agronomic parameters in paddy rice; it can be said that producers in Turkey also largely adhere to these standards.

Our findings are also consistent with these reports: while agronomic practices have become largely homogeneous across provinces, yield differences are largely due to environmental and edaphic factors, which are more difficult to manage. However, certain regional agronomic constraints should not be overlooked: Avcuoğlu's (2022) finding of the prevalence of the paddy white-tip nematode (Aphelenchoides besseyi) in Gönen indicates that biotic stress factors can be a significant cause of yield loss at the local level; meanwhile, the rich community of beneficial insects identified by Duman et al. (2013) in Karacadağ demonstrates that biological control can be a tool within integrated pest management. The study by Özçakmak et al. (2017) on HACCP practices, on the other hand, indicates that farm-to-table quality and food safety should be addressed in an integrated manner with agronomic sustainability.

Structural Significance of the Spatial Distribution

The spatial pattern depicted in Figure 9 places the study's numerical findings within a geographical context. In Türkiye, paddy rice production takes place over a small proportion of the country's surface area – estimated to be less than 3%; this indicates that paddy rice production exhibits a 'basin-dependent' pattern, concentrated within three defined basins, rather than being a 'national' production item. This geographical concentration confirms the decisive role of irrigation infrastructure and water access in determining yield (Sokat, 2023; Nie et al., 2024); it also constitutes the geographical reflection of the fact that soil moisture was ranked highest in the Random Forest analysis conducted in this study.

As a result of this spatial asymmetry, the structural supply gap in Türkiye's paddy–rice market cannot be sustainably closed. Çelik and Emeksiz (2019)

demonstrated that, as of 2017, national production could meet only approximately 540,000 tonnes of the 750,000 tonnes of domestic consumption, with the remaining gap being filled by imports. Two parallel strategies stand out to close this gap: (i) further increasing the yield potential of existing high-yield basins (Thrace, Bafra–Çarşamba) through modern irrigation, variety improvement and site-specific nutrient management; and (ii) integrated planning of organic matter, pH and irrigation management to reduce the yield gap in low-yield provinces in the Çukurova–GAP region where water access is available through GAP infrastructure (Kahraman et al., 2021; Öktem, 2016; Saberali & Darzi-Naftchali, 2024).

Methodological Assessment

In this study, the high performance achieved by the RF algorithm, with an accuracy of 93.47% and an AUC of 0.9764, is in line with the literature on machine learning–based crop yield modelling. Jeong et al. (2016) reported that they achieved $R^2 > 0.90$ values for paddy yield estimation using a similar RF approach; meanwhile, Chen et al. (2024) achieved an R^2 value of 0.87 for RF regression. Ibrahim et al. (2023) reported accuracy in the range of 85%–92% in their approach, which combined PlanetScope high-resolution satellite imagery with various ML algorithms; Abarna and Kesavaraja (2024), on the other hand, achieved accuracy values above 95% using a CNN–RNN hybrid deep learning architecture. A recent review by Liu et al. (2025) highlights that RF consistently achieves accuracy levels above 90% in multivariate classification problems. The fact that the OOB error (6.3%) is consistent with the test error (6.53%) is an important indicator of the model's robustness.

However, it is also important to consider certain methodological limitations of the study. First, the dataset is limited to 160 observation points representing four provinces and does not cover all paddy production regions in Türkiye. Secondly, the 14 parameters used in the study are based on field measurements and secondary literature, and enhancing them with remote sensing products of higher spatial and temporal resolution (e.g., Sentinel-2 NDVI series and PlanetScope images; see Ibrahim et al., 2023) could yield more robust results. Ibrahim et al., 2023) could yield more robust results. Thirdly, the study adopted a cross-sectional approach, and examining the long-term effects of climate change using panel data or time-series models constitutes a productive research agenda for future studies (Cai et al., 2023; Huang et al., 2024). Fourth, although the assumptions of the ANOVA used in the study were met at the level of the four provinces, it should be borne in mind that non-parametric alternatives (Kruskal–Wallis) or hierarchical mixed models may be more appropriate when more provinces are included.

Conclusion and Recommendations

In this study, 14 environmental, edaphic and agronomic parameters affecting paddy (Oryza sativa L.) yield in Türkiye were analysed using Random Forest classification and one-way ANOVA on a dataset of 160 observations compiled from provinces with high yield (Kırıkkale), medium yield (Balıkesir), low yield (Diyarbakır) and the lowest yield (Şanlıurfa). The main findings of the study can be summarised as follows:

First, the Random Forest model predicted paddy yield classes with a high level of accuracy, achieving values of 93.47%, Cohen's kappa of 0.912, and ROC-AUC of 0.9764. The consistency of the out-of-bag error with the test error supports the generalisability of the model.

Secondly, the Gini-based variable importance analysis revealed that the five most influential parameters on paddy yield were, in order, soil moisture, organic matter content, soil pH, precipitation and temperature. These five parameters account for approximately 72% of the total relative importance, which is consistent with PCA and the international literature.

Third, the ANOVA results confirmed the RF importance ranking; a significant difference ($p < 0.001$) was found for all five parameters with the highest importance. This consistency indicates both methodological robustness and the statistical reliability of the findings.

Fourth, the fact that agronomic practices (plant population, fertiliser rate, weed control frequency) do not result in significant differences between provinces indicates that paddy growers in Türkiye largely comply with agronomic recommendations, and that yield variations are primarily due to environmental and edaphic factors.

Fifth, Figure 9 (provincial paddy production density map), newly added to the study, provides a spatial visualisation of the fact that paddy production in Türkiye is a basin-dependent agricultural sector concentrated in the Thrace–Marmara, Bafra–Çarşamba and, to a limited extent, Çukurova–GAP basins. This pattern of geographical concentration is a geographical reflection of the fact that the Random Forest analysis placed soil moisture in first place.

In terms of policy and practice, the study puts forward five key recommendations: (i) improving irrigation infrastructure and expanding the use of pressurised/drip irrigation systems in low-yielding provinces such as Diyarbakır and Şanlıurfa; (ii) increasing organic matter input (green manure, paddy straw return, compost) in these regions; (iii) planning region-specific acidifier or gypsum applications to balance soil pH; (iv) developing data-driven decision support systems at the national level and regularly analysing TURKSTAT data using machine learning–based model updates; and (v) adapting site-specific nutrient management principles to Turkey's paddy production basins.

The proposed research agenda for future studies encompasses the following topics: higher-resolution analyses covering all paddy-producing provinces in Türkiye (26 provinces) and enhanced with Sentinel-1/2 and PlanetScope remote sensing data; combining paddy yield projections under climate change scenarios with CMIP6 models; comparative testing of more advanced ensemble and neural network architectures, such as XGBoost, LightGBM and deep learning (CNN, LSTM); and the application of explainable artificial intelligence techniques, such as SHAP (SHapley Additive exPlanations), to interpret the interactions of variables.

In conclusion, this study fills a gap in the literature on multivariate statistical analyses of paddy yield in Türkiye

and provides outputs that are actionable from both a methodological and a regional policy perspective. The combined use of Random Forest and ANOVA provides a methodological framework that can serve as a reference not only for the four provinces covered by this study but also for agricultural yield studies in Türkiye in general.

Declarations

Ethics Approval Certificate

Ethics Committee approval is not required for this study.

Author Contribution Statement

Kerim Karadağ: Established the necessary technical infrastructure and developed the method and methodology for analysing the data. Developed the decision on which analysis software would analyse the data and how. Employed machine learning.

Mehmet Özcanlı: Carried out data collection, research, selected the cities to be used for analysis (case studies), determined the geographical parameters for comparing the case studies, and classified the relevant data.

The authors contributed equally to the interpretations.

Funding Statement

This study was not funded by any institution/organisation.

Conflict of Interest

The authors do not declare any conflicts of interest.

Acknowledgements

We would like to express our gratitude to the Department of Computer Engineering at Haran University for providing the advanced computer technologies and technical support required for the analysis.

References

Abarna, S., & Kesavaraja, D. (2024). Dynamic parameter calibration based deep network for paddy yield prediction. Computers and Electronics in Agriculture (in press). https://doi.org/10.1016/j.eja.2025.127929

Avcuoğlu, Ö. (2022). Çeltik beyaz uç nematodu Aphelenchoides besseyi Christie, 1942 (Aphelenchida: Aphelenchoididae)'nun Gönen (Balıkesir) ilçesi çeltik ekim alanlarındaki yaygınlığı [Yüksek lisans tezi]. Çanakkale Onsekiz Mart Üniversitesi, Lisansüstü Eğitim Enstitüsü, Bitki Koruma Anabilim Dalı. https://acikerisim.comu.edu.tr/items/bc185447-5dc6-4d9f-9f3f-e2f9ffe45dcb

Bai, W., Zhang, H., Liu, B., Wu, Y., & Song, Z. (2024). Soil pH determines nitrogen effects on methane emissions from rice paddies. Global Change Biology, 30(11), e17556. https://doi.org/10.1111/gcb.17556

Beşer, N. (2013). Rice booms in Turkey. Rice Today, 12(3), 36–37. International Rice Research Institute. https://ricetoday.irri.org/rice-booms-in-turkey/

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

Cai, S., Zhao, X., Pittelkow, C. M., Fan, M., Zhang, X., & Yan, X. (2023). Optimal nitrogen rate strategy for sustainable rice production in China. Nature, 615(7950), 73–79. https://doi.org/10.1038/s41586-022-05678-x

Çavuş, C. Z., & Kırmızı Erdal, C. (2020). Umurbey Çayı Havzası'nda (Çanakkale-Lapseki) tarımı etkileyen faktörlerin coğrafi analizi. Journal of Awareness, 5(4), 571–600. https://doi.org/10.26809/joa.5.041

Çelik, H., & Emeksiz, F. (2019). Türkiye'de pirinç üretimi, tüketimi ve pazarlaması. Çukurova Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 37(1), 124–135. https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=whUQrUweVzzjCp5IqASkpQ&no=kAy4Czy7djMk1sBanVMsPQ

Chen, Y., Wang, J., Zhang, L., & Liu, H. (2024). Rice yield estimation using multi-temporal remote sensing data and machine learning: A case study of Jiangsu, China. Agriculture, 14(4), 638. https://doi.org/10.3390/agriculture14040638

Cheng, M., Jiao, X., Shi, L., Peñuelas, J., Kumar, L., Nie, C., Wu, T., Liu, K., Wu, W., & Jin, X. (2022). High-resolution crop yield and water productivity dataset generated using random forest and remote sensing. Scientific Data, 9, 641. https://doi.org/10.1038/s41597-022-01761-0

Climate Change in Turkey. (2024). Effects of climate change on agriculture in Turkey. https://climatechangeinturkey.com/effects-of-climate-change-agriculture.html

Demir, M. (2021). Kars ilinde tarım faaliyetlerinin coğrafi esasları. Doğu Coğrafya Dergisi, 26(45), 77–106. https://doi.org/10.17295/ataunidcd.894670

Dobermann, A., Witt, C., & Dawe, D. (Eds.). (2004). Increasing productivity of intensive rice systems through site-specific nutrient management. International Rice Research Institute (IRRI) and Science Publishers. http://books.irri.org/9712201872_content.pdf

Duman, M., Mutlu, Ç., Büyük, M., & Karaca, V. (2013). Karacadağ çeltik ekim alanlarında bulunan faydalı böcek, örümcek ve polinatör türlerin belirlenmesi. Türkiye Biyolojik Mücadele Dergisi, 4(1), 53–64. https://dergipark.org.tr/tr/download/article-file/208583

Elavarasan, D., Vincent, D. R., Sharma, V., Zomaya, A. Y., & Srinivasan, K. (2018). Forecasting yield by integrating agrarian factors and machine learning models: A survey. Computers and Electronics in Agriculture, 155, 257–282. https://doi.org/10.1016/j.compag.2018.10.024

Feng, L., Zhang, Y., Wang, S., Liu, B., & Li, H. (2021). Rice paddy soils are a quantitatively important carbon store according to a global synthesis. Communications Earth & Environment, 2, 154. https://doi.org/10.1038/s43247-021-00229-0

Gaytancıoğlu, O. (1997). Türkiye'de çeltik (paddy) üretim, fiyat ve pazarlama politikalarının değerlendirilmesi [Doktora tezi]. Trakya Üniversitesi, Fen Bilimleri Enstitüsü. https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=kb4hN9MEHQXeuJFAWVJvCA&no=kb4hN9MEHQXeuJFAWVJvCA

Geisseler, D., Linquist, B. A., & Lazicki, P. A. (2017). Effect of fertilization on soil microorganisms in paddy rice systems—A meta-analysis. Soil Biology and Biochemistry, 115, 452–460. https://doi.org/10.1016/j.soilbio.2017.09.018

Guo, W., Ma, J., Song, X., Yang, B., Zhu, X., & Chen, F. (2024). Land consolidation of dryland-to-paddy conversion in China: Analysis of spatial patterns, impacts on cropland ecosystem services, and adaptive management strategies. Environmental Impact Assessment Review, 108, 107594. https://doi.org/10.1016/j.eiar.2024.107594

Horie, T., Karahara, I., & Katsuhara, M. (2012). Salinity tolerance mechanisms in glycophytes: An overview with the central focus on rice plants. Rice, 5, 11. https://doi.org/10.1186/1939-8433-5-11

Huang, B., Ji, L., & Wang, F. (2024). Prediction of greenhouse gas emissions and their influencing factors in global paddy fields under future climate change. Environmental Research, 254, 119187. https://doi.org/10.1016/j.eja.2026.128081Get rights and content

Ibrahim, S., Balzter, H., & Ozigis, M. S. (2023). Estimating paddy rice yield using PlanetScope imagery and machine learning. International Journal of Applied Earth Observation and Geoinformation, 123, 103447. DOI: 10.1016/j.atech.2025.101447

Jagadish, S. V. K., Craufurd, P. Q., & Wheeler, T. R. (2007). High temperature stress and spikelet fertility in rice (Oryza sativa L.). Journal of Experimental Botany, 58(7), 1627–1635. https://doi.org/10.1093/jxb/erm003

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K.-M., Gerber, J. S., Reddy, V. R., & Kim, S.-H. (2016). Random forests for global and regional crop yield predictions. PLoS ONE, 11(6), e0156571. https://doi.org/10.1371/journal.pone.0156571

Kadıoğlu, Y. (2008). Uşak'ta haşhaş tarımının coğrafi özellikleri. Doğu Coğrafya Dergisi, 13(18), 165–188. https://dergipark.org.tr/tr/download/article-file/26889#:~:text=kuzey%2Dg%C3%BCney%20y%C3%B6n%C3%BCndeki%20geni%C5%9Fli%C4%9Fi%205,(Kartal%20Tepe)%20y%C3%BCkseltiye%20sahiptir.

Kahraman, Ş., Atakul, Ş., & Kılınç, S. (2019). Diyarbakır yöresinde çeltik tarımının yapısal durumu. Bahri Dağdaş Bitkisel Araştırma Dergisi, 8(1), 81–90. https://izlik.org/JA22PJ44MJ

Kahraman, Ş., Atakul, Ş., Kılınç, S., Alp, A., Duman, M., Baran, B., & Özaslan, C. (2021). Çeltik yetiştiriciliğinde organik tarım olanaklarının araştırılması ve konvansiyonel tarım ile karşılaştırılması. Toprak Bilimi ve Bitki Besleme Dergisi, 9(2), 54–61. https://doi.org/10.33409/tbbbd.970237

Kahraman, Ş., Atakul, Ş., Kılınç, S., & Alp, A. (2022). Karacadağ koşullarında konvansiyonel ve organik çeltik yetiştiriciliğinin araştırılması. Özal Tarım ve Gıda Bilimleri Dergisi, 2(1), 1–7. https://izlik.org/JA99RL26AD

Karabacak, K. (2018). Türkiye'de tütün tarımı ve coğrafi dağılışı. Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi, 58(1), 245–278. https://doi.org/10.1501/Cogbil_0000000179

Kong, X., Wu, C., & Yang, Z. (2023). Fertilization regimes and the nitrification process in paddy soils: Lessons for agricultural sustainability from a meta-analysis. Pedosphere, 33(4), 532–544. https://doi.org/10.1016/j.pedsph.2023.01.004

Kumar, R., Singh, M. P., Kumar, P., & Singh, J. P. (2022). Crop yield estimation model for Iowa using remote sensing and surface parameters. International Journal of Applied Earth Observation and Geoinformation, 107, 102695. https://doi.org/10.1016/j.jag.2022.102695

Lafitte, H. R., Yongsheng, G., Yan, S., & Li, Z.-K. (2007). Whole plant responses, key processes, and adaptation to drought stress: The case of rice. Journal of Experimental Botany, 58(2), 169–175. https://doi.org/10.1093/jxb/erl101

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R News, 2(3), 18–22.

Liu, B., Huang, L., Jiang, X., Liu, Y., Huang, G., Yang, C., Liang, Y., Cai, J., Zhu, G., & Kong, Q. (2024). Quantitative evaluation and mechanism analysis of soil chemical factors affecting rice yield in saline-sodic paddy fields. Science of the Total Environment, 929, 172584. https://doi.org/10.1016/j.scitotenv.2024.172584

Liu, J., Zhang, H., & Wang, Y. (2025). A comparative study of machine learning models in predicting crop yield. Discover Agriculture, 3(1), 77. https://doi.org/10.1007/s44279-025-00335-z

Lobell, D. B., & Field, C. B. (2007). Global scale climate–crop yield relationships and the impacts of recent warming. Environmental Research Letters, 2(1), 014002. https://doi.org/10.1088/1748-9326/2/1/014002

Montgomery, D. C. (2017). Design and analysis of experiments (9th ed.). John Wiley & Sons.

Munns, R., & Tester, M. (2008). Mechanisms of salinity tolerance. Annual Review of Plant Biology, 59, 651–681. https://doi.org/10.1146/annurev.arplant.59.032607.092911

Nie, T., Zhang, P., Chen, P., Liu, H., Jiang, L., Sun, Z., Yin, S., Wang, T., Li, T., & Zhang, Z. (2024). Impacts of climate change and water–fertilizer management on water balance dynamics in transplanting and direct-seeded paddy fields. Agricultural Water Management, 295, 108761. https://doi.org/10.1016/j.csag.2025.100070

Öktem, A. (2016). Şanlıurfa koşullarında Karacadağ çeltiği (Oryza sativa L.) yetiştiriciliği üzerine anket çalışması. Harran Tarım ve Gıda Bilimleri Dergisi, 20(2), 110–124. https://doi.org/10.19159/tutad.04360

Özçakmak, S., Gökçek, M., & Gül, O. (2017). Pirinç üretiminde HACCP sisteminin uygulaması. Celal Bayar Üniversitesi Fen Bilimleri Dergisi, 13(1), 259–273. https://doi.org/10.18466/cbayarfbe.303668

Özçelebi, M. A., & Yılmaz, C. (2020). Samsun'da çeltik üretimi ve pirinç imalat sanayi. Kesit Akademi Dergisi, 6(24), 324–349. https://doi.org/10.29228/kesit.45864

Özer, M., & Tuncel, N. B. (2016). Pirinç ve pirinç yan ürünlerinin glutensiz tahıl ürünlerinde kullanımı. Çanakkale Onsekiz Mart Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 2(2), 29–44. https://doi.org/10.28979/comufbed.277913

Özşahin, E. (2008). Gönen ovasında pirinç tarımı. Fırat Üniversitesi Sosyal Bilimler Dergisi, 18(2), 49–70. https://izlik.org/JA78PR63CL

Peng, S., Huang, J., Sheehy, J. E., Laza, R. C., Visperas, R. M., Zhong, X., Centeno, G. S., Khush, G. S., & Cassman, K. G. (2004). Rice yields decline with higher night temperature from global warming. Proceedings of the National Academy of Sciences, 101(27), 9971–9975. https://doi.org/10.1073/pnas.0403720101

Ray, D. K., Gerber, J. S., MacDonald, G. K., & West, P. C. (2015). Climate variation explains a third of global crop yield variability. Nature Communications, 6, 5989. https://doi.org/10.1038/ncomms6989

Saberali, S. F., & Darzi-Naftchali, A. (2024). Yield gap analysis and the relative importance of factors explaining yield variability in paddy fields. European Journal of Agronomy, 156, 127172. https://doi.org/10.1016/j.eja.2024.127172

Shi, W., Muthurajan, R., Rahman, H., Selvam, J., Peng, S., Zou, Y., & Jagadish, K. S. V. (2013). Source–sink dynamics and proteomic reprogramming under elevated night temperature and their impact on rice yield and grain quality. New Phytologist, 197(3), 825–837. https://doi.org/10.1111/nph.12088

Sokat, Y. (2023). Çeltik üretiminde farklı sulama sistemlerinin yabancı ot türlerine etkisi. Turkish Journal of Weed Science, 26(3), 243–252. https://izlik.org/JA24WC36JR

Sürek, H. (2002). Rice production and research in Turkey. Food and Agriculture Organization of the United Nations (FAO). https://www.fao.org/4/w8595t/w8595t04.htm

Sürek, H., & Beşer, N. (2004). The economic situation of rice in Turkey. Cahiers Options Méditerranéennes, 50, 49–57. https://om.ciheam.org/ressources/om/pdf/c50/03400008.pdf

Taban, S., Alpaslan, M., Hashemi, A. G., & Eken, D. (1997). Orta Anadolu'da çeltik tarımı yapılan toprakların bazı fiziksel ve kimyasal özellikleri. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 3(3), 457–466. https://izlik.org/JA38EA99LA

Tang, Q., Cotton, A., Wei, Z., Xia, Y., Daniell, T., & Yan, X. (2022). How does partial substitution of chemical fertiliser with organic forms increase sustainability of agricultural production? Science of the Total Environment, 803, 149933. https://doi.org/10.1016/j.scitotenv.2021.149933

Tanrıkulu, M. (2019). Tosya'da pirinç üretiminin dünü, bugünü ve geleceği. Akademik Bakış Dergisi, 71, 229–250. https://izlik.org/JA79BK77ES

Taşligil, N., & Şahin, G. (2011). Türkiye'de çeltik (Oryza sativa L.) yetiştiriciliği ve coğrafi dağılımı. Adıyaman Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 4(6), 182–203. https://izlik.org/JA74FB47BP

Tuğaç, M. G., & Torunlar, H. (2007). Tarım Arazilerinin Tarımsal Kullanım Uygunluklarının Belirlenmesi Üzerine Bir Çalışma. *Journal of Agricultural Sciences*, *13*(03), 157-165. https://izlik.org/JA78FA48TX

Tuğaç, MG, Tercan, A., Torunlar, H., Karakurt, E., & Usul, M. (2023). Yarı kurak bölgelerde GIS tabanlı çok kriterli karar analizi ve jeoistatistiksel yaklaşımla tarımsal arazi uygunluğunun değerlendirilmesi. *Toprak Çalışmaları* , *12* (1), 15-29. https://doi.org/10.21657/soilst.1328637

TÜİK. (2025). Bitkisel üretim istatistikleri: Çeltik ekili alan, üretim ve verim (2004–2024). Türkiye İstatistik Kurumu. https://data.tuik.gov.tr

Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture, 177, 105709. https://doi.org/10.1016/j.compag.2020.105709

Welch, J. R., Vincent, J. R., Auffhammer, M., Moya, P. F., Dobermann, A., & Dawe, D. (2010). Rice yields in tropical/subtropical Asia exhibit large but opposing sensitivities to minimum and maximum temperatures. Proceedings of the National Academy of Sciences, 107(33), 14562–14567. https://doi.org/10.1073/pnas.1001222107

Wu, B., Meng, J., Li, Q., Yan, N., Du, X., & Zhang, M. (2023). Rapid rice yield estimation using integrated remote sensing and meteorological data and machine learning. Remote Sensing, 15(9), 2374. https://doi.org/10.3390/rs15092374

Yang, Y., Xie, G., Wang, M., Dai, Y., Chen, W., Zhang, Y., & Zhang, T. (2024). Coupling a process-driven model with geographic information system enables the spatial prediction of cadmium in paddy soils. Environmental Pollution, 343, 123164. https://doi.org/10.1016/j.envpol.2023.123164

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., … Asseng, S. (2017). Temperature increase reduces global yields of major crops in four independent estimates. Proceedings of the National Academy of Sciences, 114(35), 9326–9331. https://doi.org/10.1073/pnas.1701762114