Rocznik Ochrona Środowiska

Volume 28 Year 2026 ISSN 2720-7501 pp. 649-657

https://doi.org/10.54740/ros.2026.043 open access

Received: May 2026 Accepted: July 2026 Published: July 2026

Agricultural Greenhouse Gas Emissions in Türkiye: A Comparison with EU-27 Countries and Machine Learning–Based Forecasting

Çağdaş Civelek

Precision Agriculture and Agricultural Robotics Department, Faculty of Agriculture and Agricultural Sciences, Niğde Ömer Halisdemir University, Niğde, Türkiye

https://orcid.org/0000-0002-3669-488X

corresponding author's email: ccivelek@ohu.edu.tr

Abstract: Greenhouse gas (GHG) emissions in Türkiye have shown a continuous increase over the past 35 years, with a total rise of 155.3%. In comparison, agricultural greenhouse gas emissions have increased by 41.8% and currently account for approximately 12% of total GHG emissions, highlighting the sector's significant contribution. Accurately determining and planning future emission levels is crucial for implementing effective mitigation strategies and aligning with European Union regulations. This study presents a comparative analysis of Türkiye's agricultural greenhouse gas emissions with those of the 27 European Union countries for the period 2015–2023 and provides forecasts for the years 2025–2030. The findings indicate that, although Türkiye ranks as the second-highest country in terms of total agricultural greenhouse gas emissions, it ranks ninth lowest in terms of emission intensity. Among the three prediction models evaluated, the Random Forest Regression model demonstrated the best performance, achieving a coefficient of determination (R^2) of 0.503, compared to 0.482 for Ridge Regression and −2.048 for Linear Regression.

Keywords: greenhouse gas emission, machine learning, agricultural greenhouse

1. Introduction

Climate change is an undeniable factor for planet Earth and harms sustainable living, especially for human beings. It is controversial that climate change mostly relies on human activity such as industry, transportation, and agricultural activity. One of the most important variables of climate change is gases released into the air due to these activities, which are called greenhouse gases. However, greenhouse gases cover less than 1% of air molecules; they trap heat reflected from Earth's surface and reflect it back to the surface, so the Earth's surface heats up.

GHG emissions are defined in the 2006 IPCC Guidelines as the agriculture, forestry, and other land use (AFAOLU) sector (IPCC, 2006). Calculation of greenhouse gas emissions for agriculture covers a complex methodology which takes inputs into account such as cropland, livestock, managed soils (Fig. 1). Required variables for calculation of greenhouse gas emissions for cropland as defined in 2006 IPCC as ground biomass, dead organic matter, soil carbon, crop residue burning and methane emissions from rice production, whereas livestock related calculation requires enteric fermentation and manure management data and managed soils requires soil management, liming and urea fertilization data (IPCC, 2006). However, gathering data related to these variables can be challenging for better prediction and for taking precautions to reduce the negative effects of greenhouse gases. To overcome this problem, predicting future GHG emission levels using today's advanced computing technologies, such as machine learning or deep learning, can be a practical solution. Several researchers have studied different machine learning models and different data repositories for future prediction of greenhouse gas emission levels.

Rokicki et al. (2020) calculated CO_2 agricultural GHG emissions for 28 European countries for years from 2004 to 2017 using several agricultural-related gases to show the relation between the economic development level of countries and agricultural production volume. They found that the highest increase in agricultural GHG emission levels occurred in Latvia, Bulgaria, Estonia, Hungary and Poland, respectively. However, the highest greenhouse gas emitters among all countries involved were France, Germany, Spain, Poland, and Italy.

Moreira et al. (2024) implemented machine learning methods to predict agricultural GHG emissions in Australia. They integrated Generalized Additive Models (GAMs) and Conditional Inference Trees (CITs) in their study and used 30 years of CO_2 equivalent data for prediction using enteric fermentation, manure management, rice cultivation, agricultural soils, field burning of agricultural residues, liming, and urea application variables to make predictions, and their model prediction accuracy had an R^2 of 0.8.

Bista et al. (2025) developed machine learning models for predicting GHG emissions in diversified semi-arid cropping systems. In their study, they used decision tree, random forest, gradient boosting, bagging regressor, and XGB booster models. They used variables such as daily CO_2 and N_2O measurements along with air temperature, precipitation, soil moisture and temperature, cover crop carbon content, and irrigation variable to predict CO_2 and N_2O emissions. They found that the random forest model had the best accuracy of all models with R^2 values of 0.68 for CO_2 and 0.56 for N_2O, respectively.

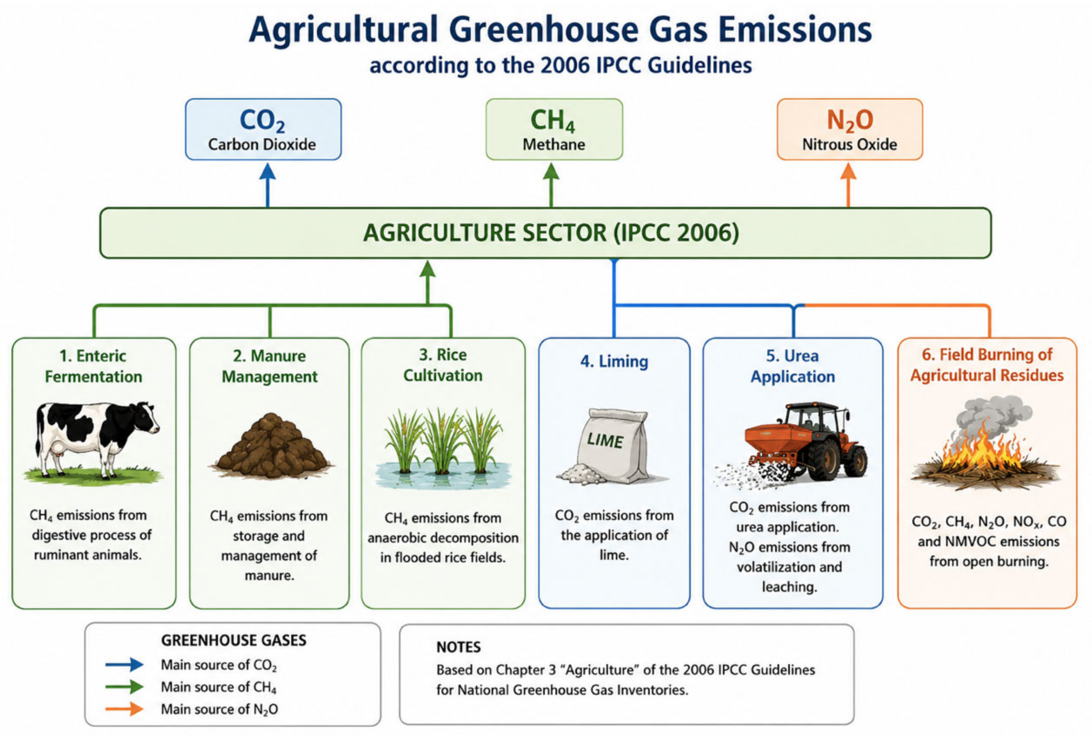


Fig. 1. GHG emissions from agriculture (IPCC, 2006)

Huan and Liu (2025) used four different machine learning models: light gradient boosting machine, extreme gradient boosting, and random forest to predict GHG emission levels resulting from rice cultivation, which is a high contributor. The developed models' prediction accuracy had R^2 of 0.75, 0.74, and 0.73, respectively.

Uyar and Uyar (2025) studied grassland and cropland changes on GHG emissions based on C and N_2O gases using remote sensing data and machine learning models. In their study, they used Landsat satellite imagery and Google Earth Engine for spatial and temporal analysis and different machine learning models, gradient boosting trees (GBT), random forest (RF), support vector machines (SVM), and classification and regression trees models (CART), to predict GHG emissions. The variables to predict C and NO_2 emissions were enhanced vegetation index, albedo, elevation, humidity, land surface temperature, normalized difference vegetation index, population density, precipitation, soil moisture, and wind speed. They found that for cropland C emissions, the GBT and RF models were the best, and for N_2O emissions, the GBT was the best-performing model. The RF model was the second-best overall.

Gathering related data to further estimation of agricultural GHG emissions is very demanding and requires complex measurement methods. For that reason, estimating future agricultural greenhouse gas emissions using newer techniques such as machine learning can provide a quick and reliable prediction method. In this article, first, the European Union's 27 countries and Türkiye's agricultural greenhouse gas emissions for the years from 2013 to 2024 were investigated using different data repositories. Secondly, to predict Türkiye's agricultural greenhouse gas emission levels, a machine learning-based method (a random forest model) was used along with two different models (linear regression and ridge regression), and the results were compared.

2. Materials and Methods

To compare agricultural GHG emission levels of EU27 countries and Türkiye, several data sources were used, such as the OECD database and EUROSTAT. Average agricultural land values were gathered using the EUROSTAT database (EUROSTAT, 2026a), and total agriculture-related GHG emission level data were gathered from the OECD database (OECD, 2026a). Using these data, specific GHG emission levels (tons CO_2/hectare) were calculated. For Türkiye, GHG emission levels for different sectors, agricultural greenhouse gas emission-related productions, and main agricultural greenhouse source gases were calculated using data gathered from the Turkish Statistical Institute.

Using 9 different inputs, Türkiye's agricultural GHG emission trend was forecasted using 3 different methods: linear regression, ridge regression, and random forest. The inputs in the developed dataset were number of cattle (TÜİK, 2026c), agricultural land (TÜİK, 2026d), agricultural ammonia (OECD, 2026b), rice production area (TÜİK, 2026d), rice production related CO_2, wheat production related CO_2, maize production related CO_2 other coarse grain production related CO_2 and sugar beet production related CO_2 (OECD, 2026c) for years from 2015 to 2024.

Three different forecasting models were used: Linear Regression, Ridge Regression, and Random Forest Regression, to predict future agricultural greenhouse gas emissions for the years between 2025–2029. The performance of the models (root mean square error, mean absolute error, mean absolute percentage error, and cross-validation values) was calculated and compared.

Linear Regression Model is calculated by using the following equation,

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+ \dots\dots\dots+\beta_nX_n+\varepsilon \tag{1}$$

where Y is the dependent variable, X_i are the independent variables, β_i is the model coefficient, and ε is the error value.

For determination of Ridge Regression model values,

$$\min_\beta \left(\sum_{i=1}^{n}\left(y_i\text{-}y_i'\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2\right) \tag{2}$$

equation was used, where λ is the regularization parameter.

For calculation of predicted values using Random Forest Regression, the equation was used,

$$Y'=\frac{1}{K}\sum_{k=1}^{K}T_k(X) \tag{3}$$

where T_k is the k number of decision tree, and K is tree number.

The Random Forest Regression model was implemented using 1,000 decision trees (n_estimators = 1000) with a fixed random seed (random_state = 42) to ensure reproducibility. The maximum tree depth was not restricted (max_depth = None), allowing trees to grow until all leaves were pure or no further split was possible. The minimum number of samples required to split an internal node was set to 2 (min_samples_split = 2), while the minimum number of samples required at a leaf node was 1 (min_samples_leaf = 1). Bootstrap sampling was enabled, and all predictor variables were considered at each split (max_features = 1.0).

For determination of model performance, Mean Squared Error (MSE), Root Mean Square Error (RMSE), and Determination Coefficient (R^2) values were also calculated.

Since the data set is very limited, LOOCV (Leave-one-out cross-validation) was performed to determine which model is the best of the three (Equation 4). In this validation test, one input is separated for testing, and the other inputs are used to train the model; the tests are repeated for each input value. For each iteration, an error is calculated, and the average error is calculated and given in the results and discussion section.

$$\text{LOOCV error}=\frac{1}{n}\sum_{i=1}^{n}\left(y_i\text{-}y'_{-i}\right)^2 \tag{4}$$

where y_i is the actual value, and y'_{-i} is the prediction of the model except i. value.

Determination of the most effective inputs on GHG emission levels: SHAP (SHapley Additive exPlanations) values were determined for each input of the models (Equation 5). The equation of the SHAP value calculation is as follows,

$$f(x)=\phi_0+\sum_{i=1}^{n}\phi_i \tag{5}$$

where ϕ_i is the contribution of the input value. If the result is positive, it has an increasing effect on greenhouse gas emission level; if it is negative, it has a decreasing effect on greenhouse gas emission level.

3. Results and Discussion

The average agricultural land for Türkiye and EU27 countries was given in Figure 2. Among these countries, Türkiye, France, Spain, Germany, and Poland were the first 5 countries with the largest agricultural land, while Cyprus, Luxembourg, Slovenia, Estonia, and Belgium were the last 5 countries with the smallest agricultural land. When agricultural greenhouse gas emissions were considered, the results were nearly the same: Germany, Türkiye, France, Italy, and Poland were the countries with the highest greenhouse gas emissions, whereas Iceland, Cyprus, Luxembourg, Latvia, and Estonia were the 5 countries with the lowest CO_2 GHG emissions.

The average GHG emissions in thousand tons CO_2-equivalent for 27 European countries and Türkiye were given for the years from 2014 to 2023 for 10 years in Figure 3. The three countries with the highest average greenhouse gas emissions were Germany, Türkiye, and France, whereas Iceland, Cyprus, and Luxembourg had the lowest average greenhouse gas emissions.

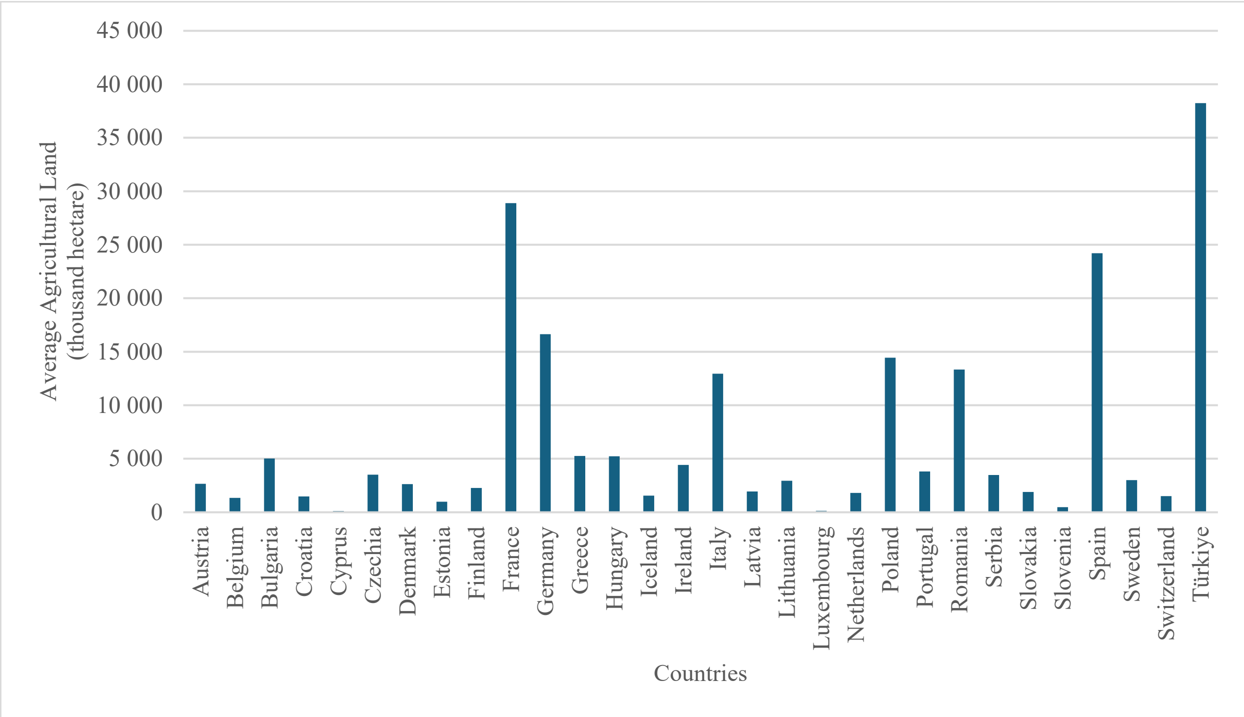


Fig. 2. Average agricultural land of EU27 countries in comparison with Türkiye from 2014 to 2023 (EUROSTAT, 2026a)

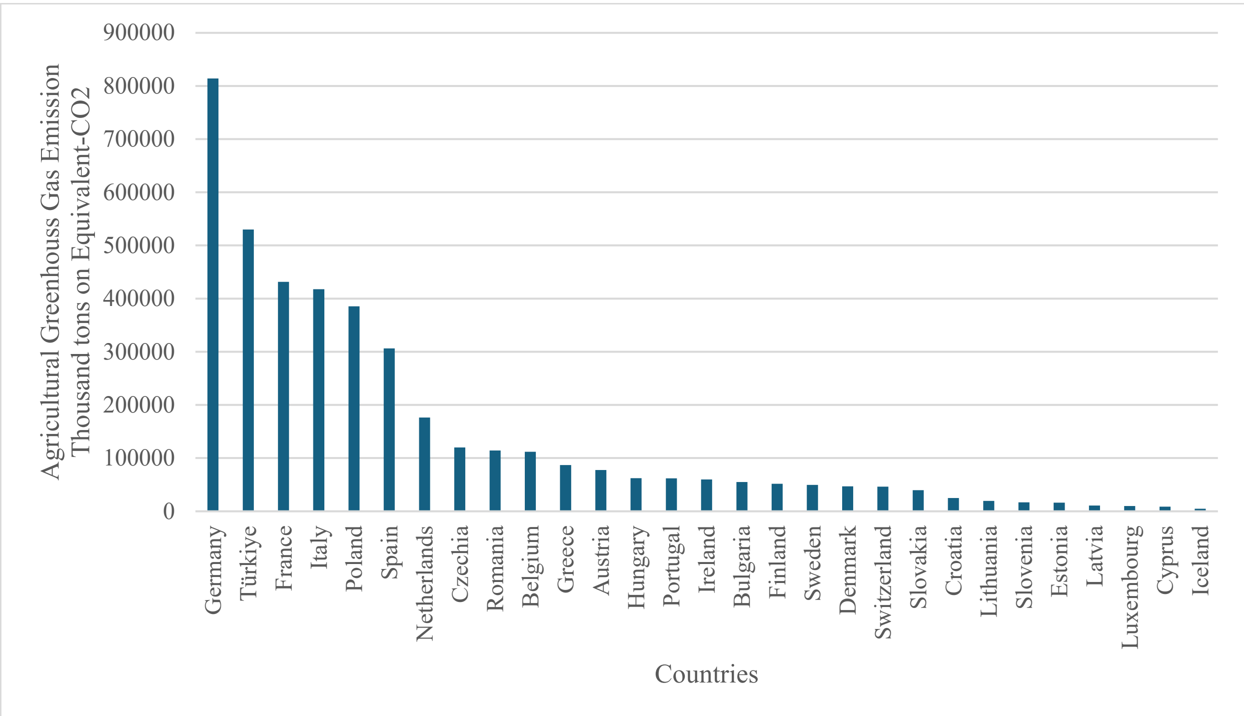


Fig. 3. Total GHG emissions from agriculture in the EU and Türkiye between 2014–2023 (OECD, 2026b)

However, for better comparison of countries for agricultural GHG emission levels, equivalent CO_2 in tons per hectare of agricultural land should be calculated, which can be defined as a specific greenhouse gas emission level. The results for this calculation are given in Table 1, and the comparison is shown in Figure 4.

According to these results, the countries with the highest specific GHG emission levels were found to be the Netherlands, Belgium, Luxembourg, Cyprus, and Germany, respectively, whereas Iceland, Latvia, Lithuania, Romania, and Bulgaria were the countries with the lowest specific greenhouse gas emission levels. It is clear from the green bars in Figure 4 that, in particular, the Netherlands, Belgium, and Germany had the highest specific greenhouse gas emission levels, despite having the smallest agricultural land. It can be stated that in those countries, agricultural production, not only plant production but also animal production, is very dense. However, Cyprus, having the smallest agricultural land among these countries, had very high specific greenhouse gas emissions. This can be attributed to heavy animal production.

For Türkiye, however, although it is the second country with the highest agricultural area, its specific GHG emission level was found to be very low, at 14 tons CO_2/hectare.

Table 1. Greenhouse gas emissions per agricultural area for EU27 countries and Türkiye between 2014–2023

Countries	Agricultural GHG emission rate (thousand tons of CO_2) (OECD, 2026a)	Agricultural Land (thousand hectares) (EUROSTAT 2026a)	CO_2/Agricultural land (tons CO_2/hectare)
Iceland	4779	1556	3
Latvia	10667	1940	5
Lithuania	19322	2943	7
Romania	114184	13335	9
Bulgaria	54993	5023	11
Hungary	62099	5226	12
Spain	306211	24212	13
Ireland	59629	4427	13
Türkiye	529979	38231	14
France	431364	28894	15
Portugal	62022	3818	16
Estonia	16260	987	16
Greece	86756	5268	16
Sweden	49618	3008	16
Croatia	24748	1475	17
Denmark	46940	2628	18
Slovakia	39536	1895	21
Finland	51742	2270	23
Poland	385565	14439	27
Austria	77457	2654	29
Switzerland	46183	1509	31
Italy	417727	12943	32
Czechia	119997	3518	34
Slovenia	16617	480	35
Germany	814103	16648	49
Cyprus	8459	122	69
Luxembourg	9732	132	74
Belgium	111863	1351	83
Netherlands	176281	1814	97

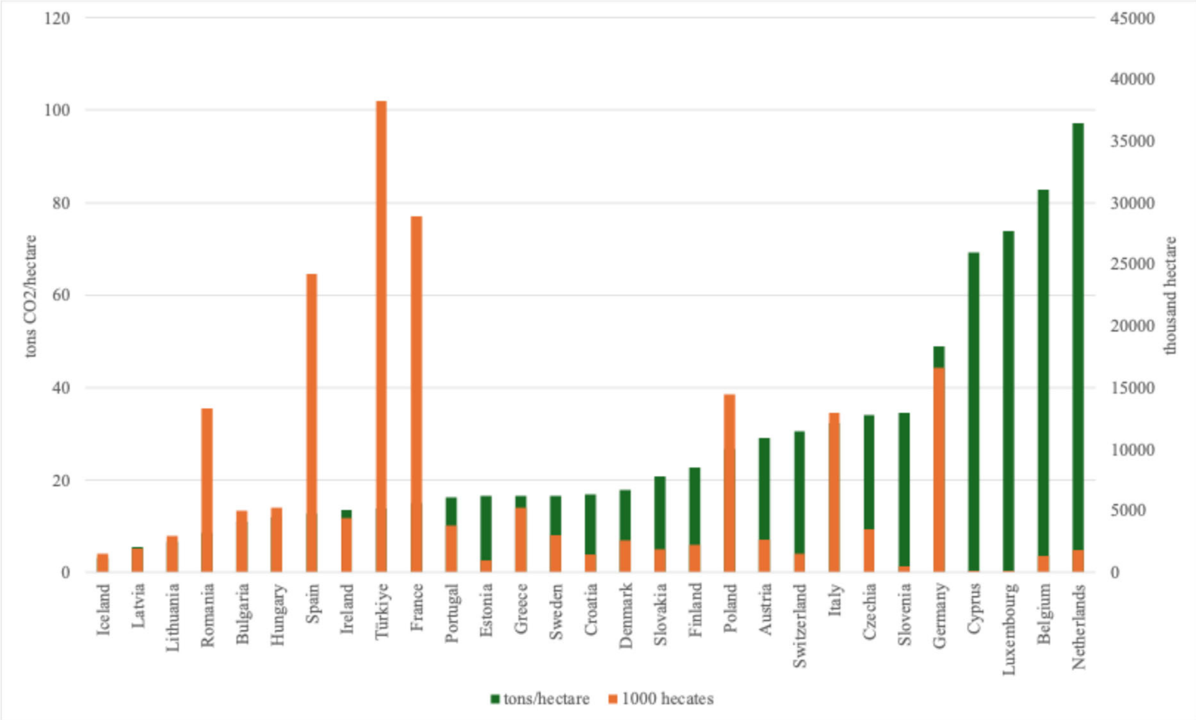
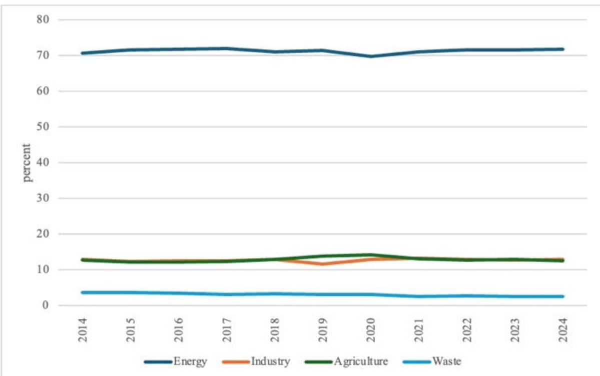
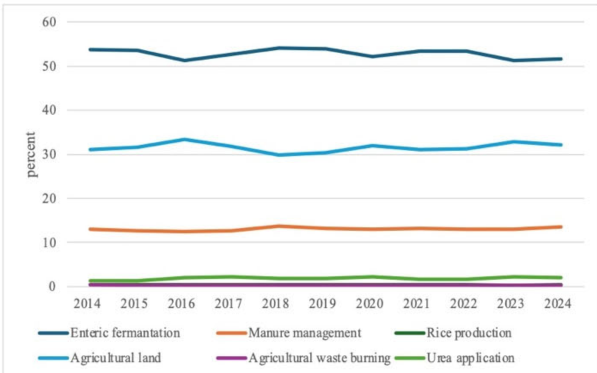


Fig. 4. Specific GHG emission level for agricultural land of EU27 countries and Türkiye between 2014–2023

To better understand the effect of agricultural GHG emissions on a country's general greenhouse gas emissions levels, it should be determined sector by sector. For Türkiye, greenhouse gas emission levels data by sector were taken into consideration using the data repository of the Turkish Statistical Institute, and the results are given in Figure 5a (TÜİK, 2026a). According to the analysis, agriculture was found to be the second sector, after industry, affecting greenhouse gas emission levels for the years between 2014 and 2023, and the emission levels were found to be steady for all sectors. Also, agricultural GHG emissions were determined based on the source of the gases. Analysis showed that the main sources of agricultural gases were enteric fermentation (around 50%), agricultural land (around 30%), and manure management (around 10%), as shown in Figure 5b (TÜİK, 2026b).



a



b

Fig. 5. a. CO_2 equivalent greenhouse gas emissions for different sectors in Türkiye (TÜİK, 2026a); b. Greenhouse gas emissions by different sectors of agricultural production (TÜİK, 2026b)

After these analyses and investigations, several variables were selected as input data for future prediction of Türkiye's agricultural GHG emissions, based on IPCC 2006 directives, and the retroactive record for years 2015 to 2024 was used to develop prediction models (Table 2). The main records were agricultural land in Türkiye, the amount of agricultural ammonia produced, rice production area, main grain production-related CO2 production, and total agricultural GHG emission levels. These data were gathered from TÜİK's data repository, and three different models were developed for prediction: linear regression, ridge regression, and random forest regression.

Table 2. Data used to predict future agricultural GHG emission levels

Years	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Number of cattle	14127837	14222228	16105025	17220903	17872331	18157971	18036117	17023791	16583005	16986259
Agricultural land (thousand hectare)	19833	19820	19713	19649	19667	19712	19972	20414	20924	21157
Agricultural ammonia (NH_3)	673208	645495	718994	756645	761035	797865	874791	874701	871990	898540
Rice production area (hectare)	115856	116056	109504	120137	126419	125398	129475	120511	112120	128903
Maize production related CO_2 (million tons CO_2)	0.85	1.02	0.94	0.79	0.90	1.09	0.84	0.82	0.90	0.91
Wheat production related CO_2 (million tons CO_2)	4.77	5.35	5.21	4.53	4.44	5.10	4.54	4.21	4.43	4.49
Other coarse grain related CO_2 (million tons CO_2)	2.92	3.56	2.94	3.06	3.46	4.41	3.16	2.76	2.99	3.10
Rice production related CO_2 (million tons CO_2)	0.97	0.98	0.92	1.01	1.06	1.06	1.08	1.08	1.13	1.15
Sugar beet production related CO_2 (million tons CO_2)	0.35	0.48	0.50	0.37	0.43	0.54	0.46	0.44	0.43	0.47
Total agricultural greenhouse gas emissions (thousand tons CO_2)	468000	481457	504993	532379	529736	517850	532133	574120	560200	598917

Using developed models, not only future prediction data for years 2025–2030 but also backwards prediction data points for years 2015–2024 were generated as shown in Figure 6. From Figure 6, it can be concluded that the backwards prediction variability of the linear prediction model was too high as compared to ridge and random forest regression models. For future predictions, linear and ridge regression models resulted in highly predicted values, whereas the random forest regression model was lower. This result is also revealed by LOOCV test results as given in Table 3. The coefficient of determination (R^2) value was found to be the highest for the Random Forest Regression model, followed by Ridge Regression. However, the R2 value for the Linear Regression Model was -0, so the prediction accuracy was not consistent, as also seen in Figure 6. Also, it can be stated that the R^2 values for the Random Forest and Ridge Regression Models were not as high as expected, so 10 years of backwards data points were not enough for the models to make good predictions. However, the best predictions gathered with the Random Forest Model were in line with the aforementioned studies.

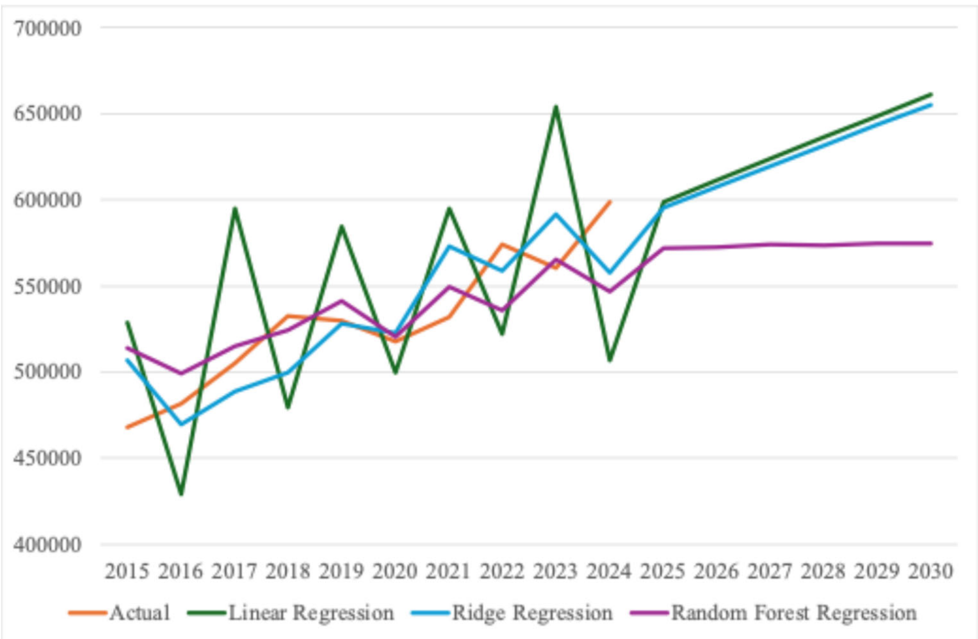


Fig. 6. Comparison of three different models' prediction of actual and future greenhouse gas emission levels

Table 3. Comparison of LOOCV values of three different models

Model	RMSE	MAE	MAPE (%)	R^2
Linear Regression	66690.119	62909.230	11.833	-2.048
Ridge Regression	27504.594	23422.703	4.408	0.482
Random Forest Regression	26918.038	20867.519	3.919	0.503

To determine the effectiveness of the Random Forest Prediction Model, SHAP analysis was also done, and mean absolute SHAP values were determined for all input variables (Fig. 7). It was determined that agricultural area size, wheat production-related CO_2, agricultural ammonia, number of cattle, and rice production-related CO_2 were the inputs with the highest SHAP values, respectively. Whereas maize production-related CO_2 and other coarse grain and sugar beet production-related CO_2 emissions did not have a huge effect on the Random Forest Prediction Model's success. So, these variables can be omitted or replaced with different variables to improve the model's success.

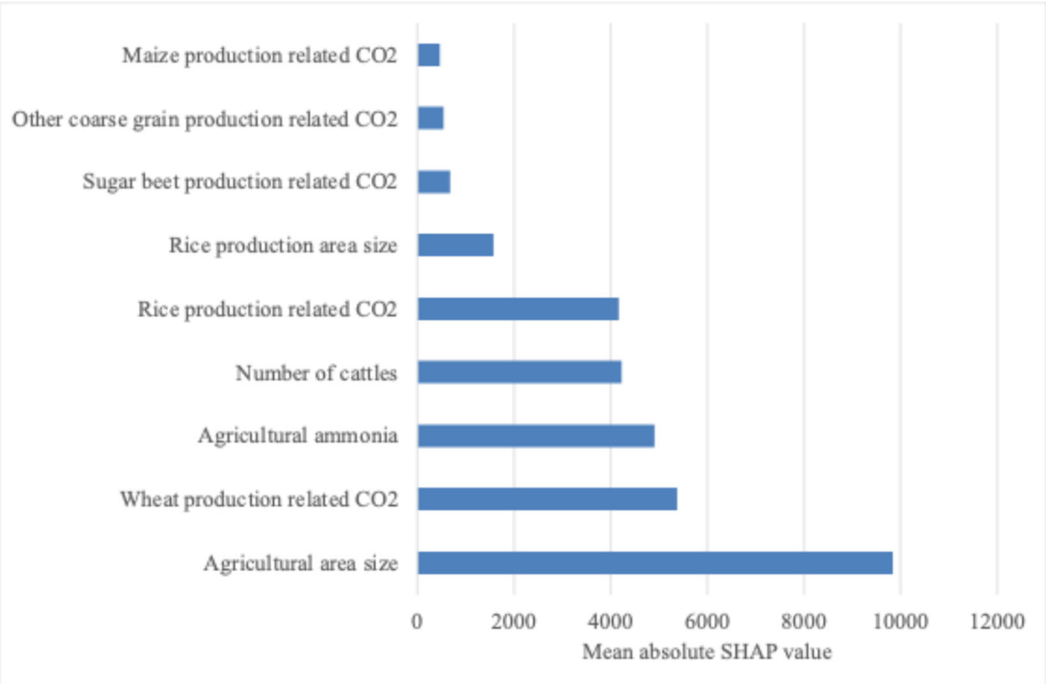


Fig. 7. SHAP (SHapley Additive exPlanations) values of inputs affecting future greenhouse gas emissions

4. Conclusion

Greenhouse gas emissions play a critical role in global warming, and agriculture is one of the major contributing sectors after energy production. Reducing agricultural GHG emissions can significantly impact a country's overall emission levels. Several factors are considered in the calculation of agricultural emissions, including agricultural land area, rice production, agricultural ammonia emissions, and enteric fermentation. However, since countries differ in their agricultural land size, comparing total agricultural GHG emissions can be misleading. Therefore, specific emission levels, expressed as tons of CO_2 per hectare, provide a more meaningful basis for comparison.

As an important agricultural producer and a candidate for European Union membership, Türkiye's compliance with EU GHG emission regulations is crucial. Although Türkiye's current average emission level is lower than the EU-27 average, forecasting future emissions is essential for effective agricultural planning and for identifying potential risks.

Advancements in computational technologies have enabled the use of various predictive methods, including artificial neural networks, machine learning, and deep learning models. Among these, Random Forest—an ensemble learning method—has demonstrated superior prediction performance compared to traditional approaches such as linear regression. However, the results showed that the Random Forest model had the highest performance; the 10-year data used for prediction is very limiting for the best results. The reason for this is that the OECD and EUROSTAT data repository had 10-year backward-recorded data. For better long-term prediction results, backward-recorded data should be used.

The results of this study also indicate that the availability of long-term historical data is essential for improving prediction accuracy. Furthermore, integrating remote sensing data—such as aerial photographs, satellite imagery, radar data, and GIS (Geographic Information Systems)—as recommended in the IPCC Guidelines (2006), can enhance the reliability of future agricultural GHG emission predictions.

References

Bista, P., Shakya, S.R., Acharya, P., & Ghimire, R. (2025). Evaluating machine learning models for greenhouse gas emissions prediction in diversified semi-arid cropping systems. *Soil Science Society of America Journal*, *89*(2). https://doi.org/10.1002/saj2.70057

Huan, S., & Liu, X. (2025). Machine learning-driven analysis of greenhouse gas emissions from rice production in major Chinese provinces: Identifying key factors and developing reduction strategies. *European Journal of Agronomy*, *164*. https://doi.org/10.1016/j.eja.2025.127536

Moreira, B.R.A., Hine, D., & Yadav, S. (2024). Integrating machine learning methods for computing greenhouse gas emissions baselines in agriculture. *Journal of Cleaner Production*, *485*. https://doi.org/10.1016/j.jclepro.2024.144416

Rokicki, T., Koszela, G., Ochnio, L., Golonko, M., Zak, A., Szczepaniuk, E.K., Szczepaniuk, H., & Perkowska, A. (2020). Greenhouse Gas Emissions by Agriculture in EU Countries. *Rocznik Ochrona Środowiska*, *22*. 809–824.

Uyar, N., & Uyar, A. (2025). Assessing climate change impacts on cropland and greenhouse gas emissions using remote sensing and machine learning. *Atmosphere*, *16*, 418. https://doi.org/10.3390/atmos16040418

EUROSTAT (2026a). *Utilized agricultural area by categories*. Accessed on 19 April 2026. https://ec.europa.eu/eurostat/databrowser/view/tag00025/default/table?lang=en&category=t_agr.t_apro.t_apro_cp

EUROSTAT (2026b). *Greenhouse gas emissions from agriculture by percent*. Accessed on 19 of April 2026. https://ec.europa.eu/eurostat/databrowser/view/tai08/default/table?lang=en

IPCC (2006). *2006 IPCC Guidelines for national greenhouse gas inventories*. Institute for Global Environmental Strategies 2108–11, Kamiyamaguchi Hayama, Kanagawa Japan, 240–0115. Accessed on 4 May 2026. https://www.ipcc-nggip.iges.or.jp/public/2006gl/vol4.html

OECD (2026a). *Greenhouse gas emissions from agriculture in CO2-equivalent tons*. Accessed on 19 of April 2026. https://data-explorer.oecd.org/vis?fs[0]=Topic%2C0%7CAgriculture%20and%20fisheries%23AGR%23&pg=0&fc=Topic&bp=true&snb=36&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_AGRI_ENV%40DF_AGGHGSEM&df[ag]=OECD.TAD.ARP&df[vs]=1.1&dq=.A.TOTGHG_GAZ....&pd=2012%2C&to[TIME_PERIOD]=false&vw=tb

OECD (2026b). *Agricultural ammonia*. accessed on 2 May 2026. https://data-explorer.oecd.org/vis?lc=en&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_AGRI_ENV%40DF_AGAMMO&df[ag]=OECD.TAD.ARP&df[vs]=1.0&dq=TUR.A.NH3AGR....&pd=1990%2C2021&to[TIME_PERIOD]=false&vw=tb

OECD. (2026c). *Agricultural production related CO_2 values*. Accessed on 2 May 2026. https://data-explorer.oecd.org/vis?fs[0]=Topic%2C1%7CAgriculture%20and%20fisheries%23AGR%23%7CAgricultural%20trade%20and%20markets%23AGR_TRD%23&pg=0&fc=Topic&bp=true&snb=5&si=2&vw=tb&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_AGR%40DF_OUTLOOK_2023_2032&df[ag]=OECD.TAD.ATM&df[vs]=1.0&dq=TUR.A.CPC_0112%2BCPC_0114T0119%2BCPC_0113%2BCPC_EX_DDG%2BCPC_0141%2BCPC_235%2BCPC_01802%2BCPC_01801%2BCPC_0111.ECOE_TOT..&pd=2014%2C2024&to[TIME_PERIOD]=false

TÜİK. (2026a). *Greenhouse gas emissions by sectors*. Accessed on 19 April 2026. https://veriportali.tuik.gov.tr/en/databrowser/tuik/categories/3/3_4/TR,DF_SERA_GAZI_EMISYONLARI,1.0

TÜİK. (2026b). *Agricultural greenhouse gas emissions by gas resource*. Accessed on 19 April 2026. https://veriportali.tuik.gov.tr/en/databrowser/tuik/categories/3/3_4/TR,DF_SERA_GAZI_EMISYONLARI,1.0

TÜİK. (2026c). *Number of cattle*. Accessed on 2 May 2026. https://biruni.tuik.gov.tr/medas/?kn=101&locale=tr

TÜİK. (2026d). *Agricultural land*. Accessed on 2 May 2026. https://biruni.tuik.gov.tr/medas/?kn=92&locale=tr