



# Future prediction of biogas potential and CH<sub>4</sub> emission with boosting algorithms: the case of cattle, small ruminant, and poultry manure from Turkey

Ihsan Pence<sup>1</sup> · Kazım Kumaş<sup>2</sup> · Melike Siseci Cesmeli<sup>1</sup> · Ali Akyüz<sup>2</sup>

Received: 26 June 2023 / Accepted: 23 February 2024 / Published online: 5 March 2024  
© The Author(s) 2024

## Abstract

Animal waste can be converted into a renewable energy source using biogas technology. This process has an impact on greenhouse gas emissions and is a sustainable source of energy for countries. It can reduce the effects of climate change and protect the planet for future generations. Tier1 and tier2 approaches are commonly used in the literature to calculate emissions factors. With boosting algorithms, this study estimated each animal category's biogas potential and CH<sub>4</sub> emissions (tier1 and tier2 approach) for 2004–2021 in all of Turkey's provinces. Two different scenarios were created in the study. For scenario-1, the years 2020–2021 were predicted using data from 2004 to 2019, while for scenario-2, the years 2022–2024 were predicted using data from 2004 to 2021. According to the scenario-1 analysis, the eXtreme Gradient Boosting Regressor (XGBR) algorithm was the most successful algorithm with an  $R^2$  of 0.9883 for animal-based biogas prediction and 0.9835 and 0.9773 for animal-based CH<sub>4</sub> emission predictions (tier1 and tier2 approaches) for the years 2020–2021. When the mean absolute percentage error was evaluated, it was found to be relatively low at 0.46%, 1.07%, and 2.78%, respectively. According to the scenario-2 analysis, the XGBR algorithm predicted the log10 values of the animal-based biogas potential of five major cities in Turkey for the year 2024, with 11.279 for Istanbul, 12.055 for Ankara, 12.309 for Izmir, 11.869 for Bursa, and 12.866 for Antalya. In the estimation of log10 values of CH<sub>4</sub> emission, the tier1 approach yielded estimates of 3.080, 3.652, 3.929, 3.411, and 3.321, respectively, while the tier2 approach yielded estimates of 1.810, 2.806, 2.757, 2.552 and 2.122, respectively.

**Keywords** Animal waste · Environmental effect · Machine learning · Methane · Regression · XGBR algorithm

## Introduction

The need for energy worldwide is increasing in line with the population, and this energy demand continues to increase with living conditions and technological developments. Energy demand has increased by 150% in the last 40 years. Energy consumption is thought to have caused an increase

of approximately 60% in air pollution. Using clean energy sources is very important to prevent such adverse effects. Greenhouse gases formed during energy production are known as the gases responsible for climate change worldwide. Since these gases' physical and chemical properties are different, their effects are also different. Due to this global situation, the Kyoto Protocol and the Climate Change Convention were signed in the United Nations Framework Convention on Climate Change (Can 2020; OECD 2015; Senocak and Guner Goren 2022).

The increase in industrial-scale livestock farming worldwide and the resulting animal waste have become significant sources of environmental pollution. Turkey is an important agricultural and livestock country. Livestock farming contributes significantly to agricultural production, rural development, and the economy. Due to Turkey's growing population and economic development, there is an increasing demand for meat and milk. However, livestock production

Responsible Editor: Ta Yeong Wu

✉ Ali Akyüz  
aliozhanakyuz@gmail.com

<sup>1</sup> Department of Software Engineering, Bucak Technology Faculty, Burdur Mehmet Akif Ersoy University, Burdur 15300, Turkey

<sup>2</sup> Bucak Emin Gulmez Vocational School of Technical Sciences, Burdur Mehmet Akif Ersoy University, Burdur 15300, Turkey

needs to be controlled. A significant amount of manure is produced in the livestock sector every day. Ensuring hygiene in the livestock sector and proper disposal of the obtained manure are crucial. The irregular disposal of livestock manure and its direct use as fertilizer in agricultural lands are two significant issues in animal manure management in Turkey (Can 2020; Erdogdu et al. 2019; Melikoglu and Menekse 2020; Şenol et al. 2021). A new regulation has been introduced to control animal manure management. This new regulation, effective in the second half of 2021, restricts the direct application of livestock manure to the soil. In addition, livestock farms are held responsible for properly storing animal manure and developing manure management plans (Chandra Manna et al. 2018; Erdogdu et al. 2019; Zaidi et al. 2018). This regulation also encourages the use of animal manure for biogas production as a management strategy. Due to certain restrictions, animal manure is not directly disposed of and is mainly stored in fertilization or storage facilities. Uncontrolled animal manure storage can lead to greenhouse gas emissions and environmental issues such as odor and hygiene problems. Animal manure that does not have proper storage facilities can indirectly contaminate surface water or groundwater. Animal manure contains various microorganisms that can risk animals and humans, causing food contamination and disease outbreaks (Font-Palma 2019; Maroušek et al. 2020; Sun et al. 2021). Therefore, implementing sustainable manure management systems that reduce the environmental risk and allow for the storage, transportation, and use of manure on farms is crucial in many aspects. The livestock and agriculture sectors contribute significantly to Turkey's greenhouse gas emissions. Therefore, it is essential to target the reduction of greenhouse gas emissions resulting from manure management and find innovative solutions to tackle this significant issue (Erdogdu et al. 2019; Şenol et al. 2021; Sun et al. 2021).

There is a growing trend toward using renewable energy sources to meet the world's energy needs while reducing environmental damage. There is a growing effort among governments, scientists, and companies to develop sustainable methods of obtaining energy and to enact laws governing the use of such methods. Several sustainable energy sources have been promoted in energy production, resulting in a gradual decrease in the use of fossil fuel-based energy. The shift in electricity generation in European Union member countries is achieved through the promotion of sustainable energy sources and the reduction of reliance on fossil fuels (Gündoğan and Koçar 2022; Karaaslan and Gezen 2022; Ocak and Acar 2021).

This year, for the first time in European Union member countries, 40% of electricity generation came from renewable energy sources, while 34% was derived from fossil fuels. The increasing utilization of renewable energy sources contributes to the transformation aimed at meeting the world's

energy demand while minimizing the impact on the ecosystem. Governments, scientists, and companies are working toward legal regulations to promote energy production and improvements through sustainable methods (Cheng et al. 2021; Erdin and Ozkaya 2019; Pence et al. 2022; Yurtkuran 2021).

This situation demonstrates the effectiveness of policies that reduce dependence on fossil fuels, transition to methods that cause less harm to the environment in energy production, and promote sustainable energy sources. Increasing renewable energy sources reduces greenhouse gas emissions, decreases environmental pollution, and combats climate change (Jones and Moore 2023; Pence et al. 2022; Senocak and Guner Goren 2022).

The anaerobic decomposition of organic materials such as manure, agricultural waste, sewage sludge, and food waste produces biogas. It primarily consists of methane (CH<sub>4</sub>) and carbon dioxide (CO<sub>2</sub>), which can be converted into heat and electricity. Biogas is known as a renewable energy source. It is widely used in Europe due to its advanced technology.

The Intergovernmental Panel on Climate Change (IPCC) Guidelines recommend two general methods for estimating CH<sub>4</sub> emissions factors: tier1 and tier2. The tier1 method uses default emissions factors to calculate emission factors, while the tier2 method uses country-specific data. Compared to tier1, tier2 is more accurate because it incorporates country-specific information (Dong et al. 2006).

As a clean, renewable energy and fuel with low environmental impact, biogas is used for various purposes such as cooking, lighting, and electricity generation (Khoshgof-tar Manesh et al. 2020). It is stated that biogas utilization can reduce greenhouse gas emissions and can be one of the most important energy sources in meeting countries' energy demands (Usack et al. 2019). Biogas, mainly consisting of 50–70% CH<sub>4</sub>, can be converted into heat and electricity. Biogas is considered a renewable energy source. Renewable energy and related conversion technologies provide an alternative to fossil fuel-derived energy, which is associated with various environmental issues (Heydari et al. 2021; Wang et al. 2021). Recently, it has been stated that photovoltaic/biomass systems are more cost-effective than renewable hybrid systems (Heydari et al. 2023).

In recent years, machine learning (ML)-based models have emerged as promising tools for predicting AD processes (Andrade Cruz et al. 2022). It is possible to estimate and determine biogas production using ML models without understanding the process mechanisms (Tufaner and Demirci 2020). Several studies have utilized ML algorithms such as artificial neural networks (ANN) and random forests (RF), or combinations thereof, applied to biogas processes (Chiu et al. 2022; Gonçalves Neto et al. 2021).

Najafi and Faizollahzadeh Ardabili (2018) studied small-scale biogas production using mushroom compost.

This study employed ANFIS (adaptive neuro-fuzzy inference system) and ANN models to predict biogas production based on independent variables. The independent variables considered in the study were the carbon-to-nitrogen (C/N) ratio, reactor temperature (T), and retention time (RT) (Najafi and Faizollahzadeh Ardabili 2018).

De Clercq et al. (2019) developed an ML model to predict biogas output based on waste input. They aimed to improve biogas production in industrial facilities by designing a graphical user interface. The ML model used in the study consisted of logistic regression, support vector machine, random forest, extreme gradient boosting, and *k*-nearest neighbor regression. According to their findings, the *k*-nearest neighbor (KNN) regression model was the most suitable method for the biogas plant, achieving an accuracy of 87% on the test set (De Clercq et al. 2019).

Stolarski et al. (2020) conducted a study on developing bioenergy technologies in Denmark, Germany, Estonia, Finland, Latvia, Lithuania, Poland, Sweden, and Norway. The research focused on assessing the potential of agricultural biomass, manure, and slurry in these countries. The study revealed that Germany and Poland have the highest potential for utilizing agricultural biomass, manure, and slurry for bioenergy production. Furthermore, it was highlighted that Germany, the leading biogas producer, accounted for 92% of all biogas plants in the studied countries (Stolarski et al. 2020).

Elmaz et al. (2020) employed ML methods to predict the outcomes of biomass gasification. Their study utilized four regression techniques: polynomial regression, support vector regression, decision tree regression, and multilayer perceptron. The results showed that the multilayer perceptron and decision tree regression outperformed the other methods regarding prediction accuracy (Elmaz et al. 2020).

Das et al. (2020) utilized farm animal population data from Bangladesh between 2005 and 2018 to estimate greenhouse gas emissions using the 2006 IPCC tier1 approach. They determined that the greenhouse gas emissions from livestock in 2018 amounted to 66.59 Gg/year CO<sub>2</sub> equivalent. The study further projected that by 2020, the emissions could reach 69.87 Gg; by 2030, 80.62 Gg; by 2040, 94.64 Gg; and by 2050, 113.10 Gg/year CO<sub>2</sub> equivalent. In 2018, the total greenhouse gas emissions were composed of enteric CH<sub>4</sub> (44%), manure CH<sub>4</sub> (3.6%), direct N<sub>2</sub>O (51.5%), and indirect N<sub>2</sub>O emissions (Das et al. 2020).

Almomani (2020) developed an ANN algorithm to model and optimize cumulative CH<sub>4</sub> production from agricultural solid waste and cow manure (Almomani 2020).

Kim et al. (2020) calculated the impact of digested biogas from organic waste on natural gas and its ability to reduce CO<sub>2</sub> emissions for two Korean wastewater treatment plants (Kim et al. 2020).

Ibidhi et al. (2021) estimated the country-specific national emission factor for CH<sub>4</sub> emissions from enteric fermentation in dairy cattle in South Korea using the 2006 IPCC approach. They calculated the emission factor for different age groups of animals. With the developed emission factor for dairy cattle, it was determined that the South Korean dairy sector has the potential to reduce greenhouse gas emissions by approximately  $97 \times 10^3$  tons of CO<sub>2</sub> equivalent per year, which corresponds to a reduction of 8% from the sector's total emissions (Ibidhi et al. 2021).

Tongwane and Moeletsi (2021) conducted a study for 2019, which determined that South Africa produced 35.37 million tons of CO<sub>2e</sub> emissions, including emissions from sources such as cattle, pasture, and grasslands. CH<sub>4</sub> emissions from enteric fermentation accounted for 64.54% of the total emissions, followed by emissions from pasture, grasslands, and savannahs at 27.66%. Regarding emissions related to fertilizer management, 4.34% of the total emissions were attributed to nitrous oxide (N<sub>2</sub>O), and 3.45% were attributed to CH<sub>4</sub> emissions (Tongwane and Moeletsi 2021).

Jeong et al. (2021) estimated the biogas production of a municipal wastewater treatment plant in South Korea with deep learning-based models. In the estimation results, the *R*<sup>2</sup> value was obtained as 0.76 (Jeong et al. 2021).

Sun et al. (2021) analyzed that China reduced its total annual greenhouse gas by 2% due to biogas production from straw and its conversion (Sun et al. 2021).

Ocak and Acar (2021) evaluated the energy production potential of Turkey's Marmara region and concluded that converting agricultural and animal wastes into biogas and then into electricity is economical (Ocak and Acar 2021).

Huo et al. (2021) estimated CO<sub>2</sub> emissions from China's agricultural biomass conversion based on life cycle assessment. They predicted the potential of agricultural biomass to replace fossil energy and reduce emissions under three scenarios, considering resource endowment and bioenergy potential of crop straw and livestock manure (Huo et al. 2021).

Ludlow et al. (2021) evaluated the potential of converting organic waste into energy using lower heating values in Chile and found that this corresponded to 3.3% of the annual energy demand (Ludlow et al. 2021).

Zubir et al. (2022) utilized livestock data from Malaysia from 2010 to 2019 to estimate greenhouse gas emissions from different animal species. In the livestock sector, poultry, pigs, non-dairy cattle, and goats were predominant. Non-dairy cattle were found to be the main contributor to CH<sub>4</sub> emissions, accounting for 73.91% of enteric fermentation emissions. Regarding CH<sub>4</sub> emissions from manure management, pigs accounted for 61.49%, while poultry accounted for 26.24%. Regarding direct N<sub>2</sub>O emissions from manure management, poultry contributed 63.25%, and non-dairy cattle accounted for 20.79%. Enteric fermentation was noted

to have the largest share in total CO<sub>2</sub> equivalent emissions, surpassing 50% (Zubir et al. 2022).

Fajobi et al. (2022) comprehensively examined the studies in the literature addressing the effects of different biomass sources used in biogas production on biogas yield with different techniques. It evaluated artificial intelligence's applicability in modeling and optimizing the anaerobic digestion process for different parameters. They used the fuzzy logic-based ANFIS method to estimate biogas yield (Fajobi et al. 2022).

Hörtenhuber et al. (2022) examined Austria's greenhouse gas emissions from livestock and the effects of livestock farming on climate for the years 1990 and 2019. It has been shown that CH<sub>4</sub> reduction from livestock reduces total CO<sub>2</sub> emissions by 16% (Hörtenhuber et al. 2022).

Senocak and Guner Goren (2023) made a 5-year prediction with the support vector machine algorithm for animal, agricultural, and municipal solid wastes, which are biomass resources in Denizli province of Turkey (Senocak and Guner Goren 2023).

Sharafi et al. (2023) measured the long-term energy efficiency of Iran's significant crops between 1970 and 2019. Greenhouse gas emissions were modeled with machine learning algorithms using 17 agricultural products in five main categories as input parameters (Sharafi et al. 2023).

Zhang et al. (2023) estimated the number of biomass resources that can be used in energy in 2020 by using specific parameters and coefficients. They also evaluated the potential to reduce CO<sub>2</sub> emissions using biomass energy depending on its life cycle (Zhang et al. 2023).

Liu et al. (2023) estimated the biogas potential of agricultural waste in Hubei Province in China and evaluated the environmental and economic impact of CO<sub>2</sub> reduction (Liu et al. 2023).

Nehra and Jain (2023) examined the estimation of animal-based biomass potential and the reduction of greenhouse gas emissions in rural Haryana, India. They stated that biomass energy production could prevent emissions of approximately 1707 to 3583 million kg/year (Nehra and Jain 2023).

Ceylan et al. (2023) developed a hybrid optimization model for Manisa, Turkey, utilizing a neuro-regression approach to determine the optimal biogas power plant location (Ceylan et al. 2023).

Heydari et al. (2023) studied the optimal design of a renewable wind/solar/biomass hybrid system for grid-independent applications in Iran by comparing the performance of genetic algorithms and particle swarm optimization. Simulation results showed that the photovoltaic/biomass system is cost-effective, and particle swarm optimization yields better results (Heydari et al. 2023).

It is possible to predict environmental quantities using boosting algorithms, an ML algorithm. By combining many weak models, these algorithms produce a single robust

model. In most cases, the real-time dataset is nonlinear. Consequently, if a model cannot accurately define the dataset values, it will become underfitted and biased. In this case, boosting algorithms are necessary to reduce bias.

In this study, Adaptive Boosting (AdaBoost), Gradient Boosting, and eXtreme Gradient Boosting (XGBoost), which are popular boosting algorithms in the literature, were used to estimate the amount of biogas and CH<sub>4</sub> emissions from animal sources. Biogas and CH<sub>4</sub> quantities were calculated for 81 provinces of Turkey based on cattle, small ruminants, and poultry numbers. To determine the theoretical biogas and CH<sub>4</sub> quantities, general and specific information about cattle, small ruminants, and poultry was used, along with data about animal age, number, breed, weight, and waste quantity. A data set was created to carry out further analysis.

With Turkey taking part in the Paris Agreement, it has plans to increase renewable energy production and reduce greenhouse gas emissions with the preparations for the 2050 climate change strategy and 2030 action plan. It also aims to develop dynamics for using artificial intelligence methods in the country. In this sense, it is crucial to accurately estimate Turkey's energy potential and emissions regarding animal husbandry-related research. Detailed examination and modeling of the livestock-based biogas potential in Turkey with boosting algorithms and energy and emission estimates based on this potential for the coming years can give researchers and policy planners ideas. In this study, biogas potential and CH<sub>4</sub> emission values in cross-validation and time series format have been converted to log<sub>10</sub> and predicted with boosting algorithms for regression analysis. Boosting algorithms were preferred because they reduce bias and make more successful predictions than classical ML algorithms.

The proposed study includes two different analyses:

- (a) The first one includes the biogas potential and CH<sub>4</sub> emission values of each province in Turkey for the years 2004–2019 and the forecasts of each of these values for the years 2020–2021 with the boosting algorithms,
- (b) The second one uses the same values for 2004–2021 and makes predictions for 2022–2024.

The novelty and contribution are (i) animal-based biogas potential and CH<sub>4</sub> emission were estimated by boosting algorithms using the unique identifier of the provinces and year information; (ii) the appropriate boosting model that can make predictions for all provinces of Turkey for the coming years has been created.

This study is presented in four sections and organized as follows: In this section, statistical information about the potential of renewable energy, animal husbandry, emissions, and biogas in Turkey is presented, and a literature

review is included. In the “[Materials and methods](#)” section, the parameters used in theoretical biogas potential and CH<sub>4</sub> emission calculations and the machine learning methods recommended for estimation are explained in detail. In the third section, the experimental results of machine learning algorithms are compared under two different scenarios, and future predictions are made. In the last section, conclusions and recommendations are made.

## Materials and methods

### Theoretical biogas potential calculation

Between 2004 and 2021, CH<sub>4</sub> production through biogas was determined for each of the 81 provinces in Turkey using the animal population data of various animal species. Each animal species was categorized separately based on age, gender, and weight. Live weight values specific to animal species and breeds were obtained from farms within the provinces to determine the amount of manure. Since no representative value was available for manure production in Turkey, the percentages of live weight values obtained from the literature were used. The percentages used were 6% for cattle, 5% for small ruminants, and 4% for poultry. The daily fresh manure values were calculated separately for each province and district based on the age and species of cattle and small ruminants and separately for poultry using these percentages. The amount of animal waste varies depending on feeding practices, climate conditions, and reproductive types. The usability coefficients for each animal species were 50% for cattle, 13% for small ruminants, and 99% for poultry. Figure 1 provides the details of the animal species along with the parameters (VS, B<sub>0</sub>, MCF, MS) used in the calculation of CH<sub>4</sub>. These parameters were utilized in the tier2 approach for CH<sub>4</sub> estimation (Avcioglu and Türker 2012; Dong et al. 2006).

If animal manure is not effectively managed and processed in a biogas production system, it can result in the uncontrolled release of CH<sub>4</sub> gas into the atmosphere. The agricultural and livestock industries are significant contributors to greenhouse gas emissions, and the improper handling of animal waste exacerbates this issue. The release of CH<sub>4</sub>, a potent greenhouse gas, further contributes to global warming and climate change. Therefore, proper collection and treatment of animal manure within biogas systems are essential for mitigating greenhouse gas emissions and promoting sustainable agricultural practices (Riaño and García-González 2015).

CH<sub>4</sub> emissions are calculated using different methods, with tier1 and tier2 being the most commonly used approaches. In tier1, a simple calculation is employed, multiplying the number of animals in each category by the

emissions factor per animal. Tier2, on the other hand, is a more advanced method implemented in most developed countries. It involves considering various parameters specific to each animal species. The Intergovernmental Panel on Climate Change (IPCC) provides assumed emission factors for each livestock category, considering the average annual temperature. These emission factors reflect the range in manure volatile solids content and the application of manure management practices in different regions. They have been evaluated based on the annual temperature for each climatic region. By utilizing tier2 methodology and considering these emission factors, a more accurate estimation of CH<sub>4</sub> emissions from livestock can be obtained. This approach allows for a more comprehensive assessment of the environmental impact of livestock farming and helps develop targeted strategies for reducing greenhouse gas emissions in the agricultural sector.

The CH<sub>4</sub> emissions were calculated using different approaches. For the tier1 approach, the formula provided in Fig. 1 and the emission factors specific to each region, as listed in Table 10.11 of IPCC-2006 (Dong et al. 2006), were utilized. This method involves multiplying the emission factor by the number of animals in each category. On the other hand, the tier2 approach employed the formulas depicted in Fig. 1, along with the parameter values specified in Table 1. This more advanced method considers additional parameters and variables associated with each animal species. Tier1 and tier2 approaches allow for a comprehensive calculation of CH<sub>4</sub> emissions. These approaches provide valuable insights into the environmental impact of livestock farming and assist in devising effective strategies for mitigating greenhouse gas emissions in the agricultural sector. The IPCC approach was used to calculate CH<sub>4</sub> emissions from enteric and manure in the dairy cow system (Baek et al. 2014) and for values from beef cattle (Chen et al. 2020). In addition, an application was made for Korea in agricultural biomass calculation (Shin et al. 2016). While the greenhouse gas detection studies carried out using the tier1 approach include Ngwabie et al. (2018), the tier2 approach was used in the studies of Herrera et al. (2021) and Basak et al. (2022). While Khan et al. (2021) examined Pakistan’s biogas production potential from animal manure (Khan et al. 2021), Şenol et al. (2021) carried out studies on calculating Turkey’s biogas potential until 2030 (Şenol et al. 2021).

### Creation of the biogas and CH<sub>4</sub> quantities data set

In this study, the biogas and CH<sub>4</sub> (tier1 and tier2 approaches) quantities for each of the 81 provinces in Turkey for the years 2004–2021 were calculated based on the data obtained from the Turkish Statistical Institute (TUIK) regarding the number of cattle, small ruminants, and poultry (TUIK 2022). The calculation used general and specific information about

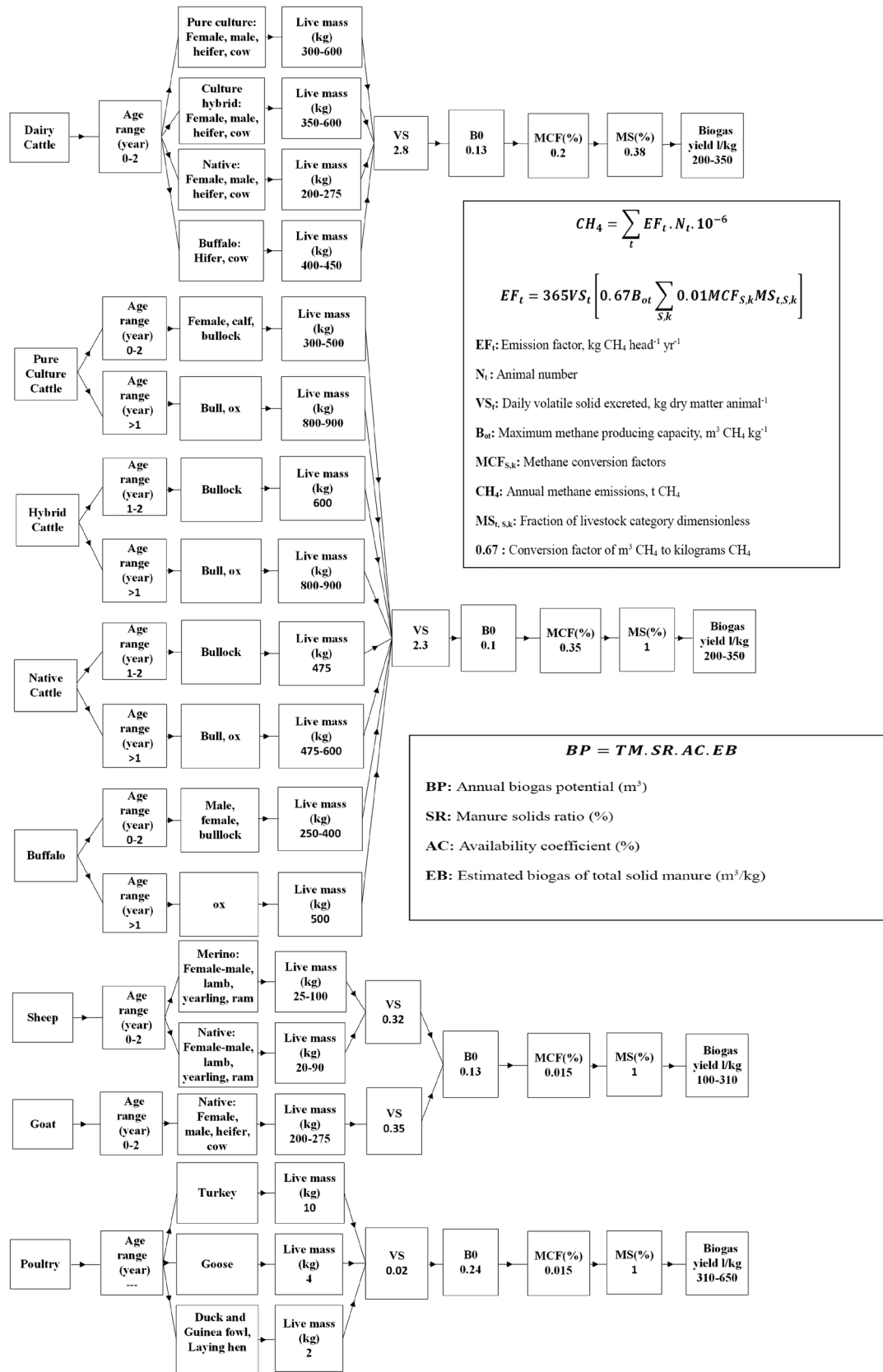


Fig. 1 Formulas and parameters for biogas and CH<sub>4</sub> emission (tier1, tier2 approaches)

**Table 1** The performance scores for modeling animal-based biogas potential for 2004–2021

Analysis	Algorithm	Train			Test/prediction		
		RMSE	MAE	$R^2$	RMSE	MAE	$R^2$
Cross-validation (2004–2019)	Gradient Boosting	0.259 ± 0.004	0.212 ± 0.003	0.8479 ± 0.0044	0.285 ± 0.016	0.232 ± 0.012	0.8134 ± 0.0239
	XGBR	0.036 ± 0.002	0.025 ± 0.001	0.9971 ± 0.0003	0.106 ± 0.013	0.071 ± 0.006	<b>0.9741 ± 0.0058</b>
	AdaBoost	0.500 ± 0.004	0.427 ± 0.004	0.4342 ± 0.0079	0.511 ± 0.015	0.434 ± 0.010	0.4040 ± 0.0447
Prediction (2020–2021)	Gradient Boosting	0.254	0.207	0.8543	0.247	0.201	0.8439
	XGBR	0.036	0.026	0.9970	0.067	0.054	<b>0.9883</b>
	AdaBoost	0.496	0.424	0.4432	0.500	0.422	0.3567

cattle, small ruminants, and poultry and data on animal age, count, breed, weight, and waste quantity to determine the theoretical biogas and CH<sub>4</sub> quantities. The biogas potential and CH<sub>4</sub> emissions (tier1 and tier2 approach) of each animal category in all provinces of Turkey for 2004–2021 were theoretically calculated, and a data set was created. While the biogas potential and CH<sub>4</sub> emissions are the target values, only the unique identifier of the provinces and year information is the input value. This way, training and prediction can be performed in a time series format. The dataset contains 1458 samples for 18 years for 81 provinces in Turkey. For scenario-1, 1296 data from 2004 to 2019 were used for training, and 162 data from 2020 to 2021 were used for testing. For scenario-2, all 1458 data from 2004 to 2021 were used for training, and the years 2022–2024 were predicted.

During regression analysis, log transformations are used to reduce the distance between data points and help develop a better model. Due to the extensive range of values and sometimes outliers included in the dataset, this may be the case. In log<sub>10</sub> transformations, base 10 is used to determine the logarithm. In addition to being more readily interpreted or checked, common logarithms are more straightforward to comprehend (Zhang et al. 2022).

This study transformed biogas potential and CH<sub>4</sub> emission values to log<sub>10</sub> because of their wide range of values in training boosting algorithms.

### Ensemble learning for prediction

Ensemble learning in ML refers to an ensemble of base learners working together to make a more accurate final prediction. As a result of high variances and biases, a single weak learner might not perform well alone. However, by combining weak learners, a strong learner can be created. It is possible to improve model performance by combining weak learners in this manner. Consequently, a problem can be solved more effectively by combining several ML algorithms. Ensemble learning algorithms train multiple models on the data and then combine their predictions to produce the final result. In regression problems, this combination is done by taking the average of the predictions. There are

various methods for model averaging, including one-step and iterative weighted parameter estimation. The objective is to increase the true conditional mean of the dependent variable provided by the predictors' prediction accuracy. This means adjusting the model's predictions under different conditions or values of the predictors to match the actual average values of the dependent variable more closely. The idea is to develop a model that can more accurately predict the conditional mean by capturing the relationships between the predictors and the dependent variable.

The total prediction error of an ML model is composed of bias and variance errors. Bias measures how far off the model's predictions are from the actual values. If the bias is high, the model does not accurately capture the relationship between the input and output variables and may need to be more complex. Variance measures how much the model's predictions change when trained on different subsets of the data. If the variance is high, the model is overfitting to the training data and may need to be simplified or trained on more data.

Ensemble learning methods such as bagging and boosting differ primarily in how they are trained. A bagging method involves training weak learners in parallel, whereas a boosting method involves training them sequentially.

The boosting technique is used in ensemble models to improve the generalization of a weak learning model, such as decision trees. Better prediction is obtained compared to the single weak learner using methods like majority voting in classification problems or a linear combination of weak learners in regression problems.

In boosting, multiple weak learners are combined to create a strong learner. A boosting algorithm differs from a bagging algorithm because it aims to reduce bias rather than variance. Boosting involves adjusting the next model's weighting based on the previous model's performance so that new subsets will contain elements that previous models had misclassified. The purpose of boosting is to improve the performance of a relatively simple classifier with a high bias rate. It is necessary to train each of the base classifiers sequentially. As a result of high bias, the model fails to capture the essential features of data

because the assumptions it makes are too basic, and boosting algorithms are used to reduce high bias.

The algorithm for boosting is as follows:

1. Initialize the dataset and assign the weights of all data points to be equal.
2. Train a weak learner on the weighted data and compute the error between predicted and actual values.
3. Increase the weights of the data points with significant errors and decrease the weights with minor errors.
4. The algorithm passes the updated weights to the next learner.
5. Steps 2–4 should be repeated until the training error is less than a predetermined threshold or for a fixed number of iterations.
6. The results from each weak learner are combined.

Boosting is a widely used technique for solving classification and regression problems. The most popular boosting algorithms in the literature are Adaboost, Gradient Boosting, and XGBoost. In this study, boosting algorithms were customized for regression analysis to estimate biogas potential and CH<sub>4</sub> emission values in cross-validation and time series format.

**Adaptive boosting**

AdaBoost is the first boosting algorithm introduced by Freund and Schapire (1997) and combines weak learners to create a strong learner. It is known as adaptive boosting because each instance receives an updated set of weights, with higher weights given to incorrectly classified instances. It combines weak classifiers iteratively trained

on incorrectly classified samples from the previous iteration into a strong classifier (Ganaie et al. 2022). This algorithm can also be used for regression problems.

As a first step in the AdaBoost algorithm,  $\omega^{(i)}$  which is the weight of each sample, is assigned an initial value of  $(1/n)$  equally. The first learner is then trained, and the weighted error rate is calculated. In Eq. (1), the weighted error rate of the  $j$ th learner is calculated.

$$r_j = \frac{\sum_{i=1}^m \omega^{(i)} \hat{y}_j^{(i)} \neq y^{(i)}}{\sum_{i=1}^m \omega^{(i)}} \tag{1}$$

In Eq. (1),  $\hat{y}_j^{(i)}$  represents the  $j$ th learner prediction for  $i$ th sample. The weights of the learners are calculated according to Eq. (2).

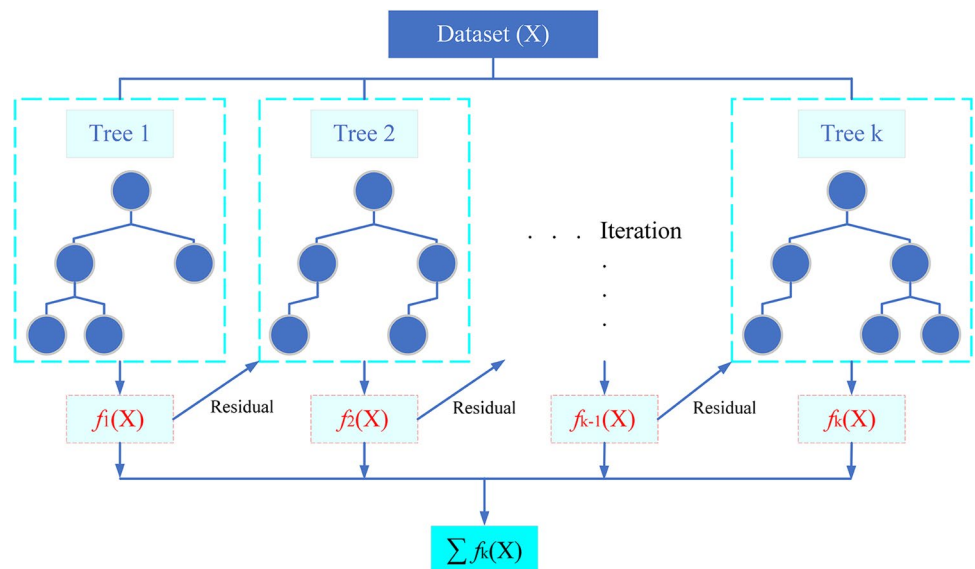
$$\alpha_j = \eta \log \frac{1 - r_j}{r_j} \tag{2}$$

In Eq. (2),  $\eta$  is the learning rate. The AdaBoost algorithm then updates the weights of the incorrectly predicted examples to speed up the learning rate, as given in Eq. (3).

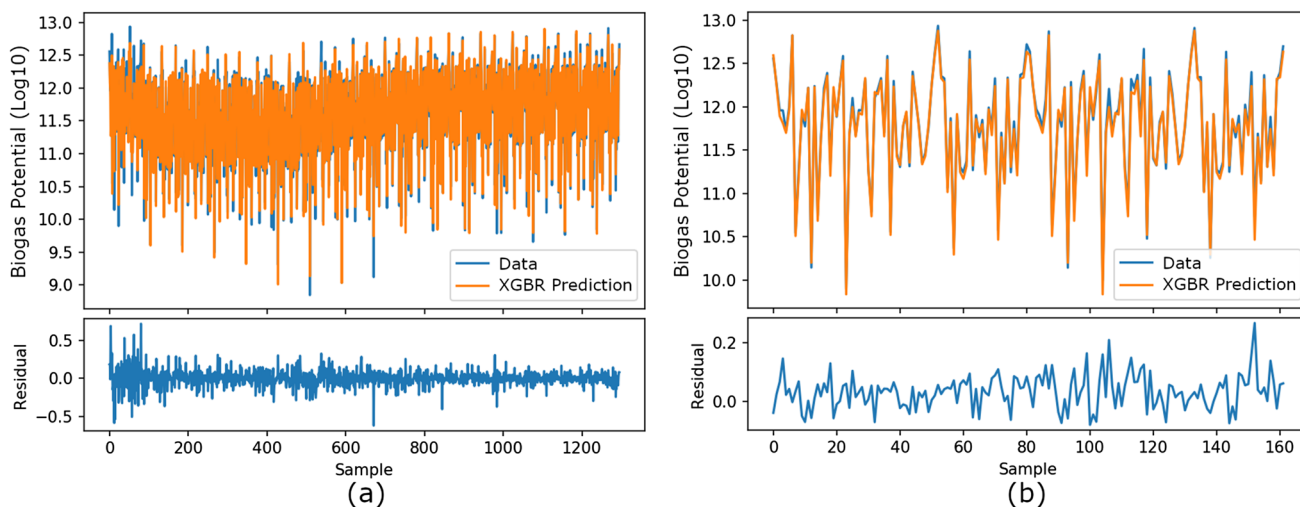
$$\omega^{(i)} = \begin{cases} i = 1, 2, \dots, m \\ \omega^{(i)}, \hat{y}_j^{(i)} = y^{(i)} \\ \omega^{(i)} e^{\alpha_j}, \hat{y}_j^{(i)} \neq y^{(i)} \end{cases} \tag{3}$$

The weights of all samples are then normalized by dividing by  $\sum_{i=1}^m \omega^{(i)}$ . Finally, a new learner is trained with the updated weights, and the process continues until the termination criterion is reached (Géron 2019).

**Fig. 2** The structure of XGBR for biogas potential and CH<sub>4</sub> emission prediction ( $f_k$ : the predicted value of each tree)







**Fig. 3** Comparison of XGBR’s predictions with the theoretical calculations of animal-based biogas potential: **a** tenfold cross-validation results and **b** prediction results for 2020–2021

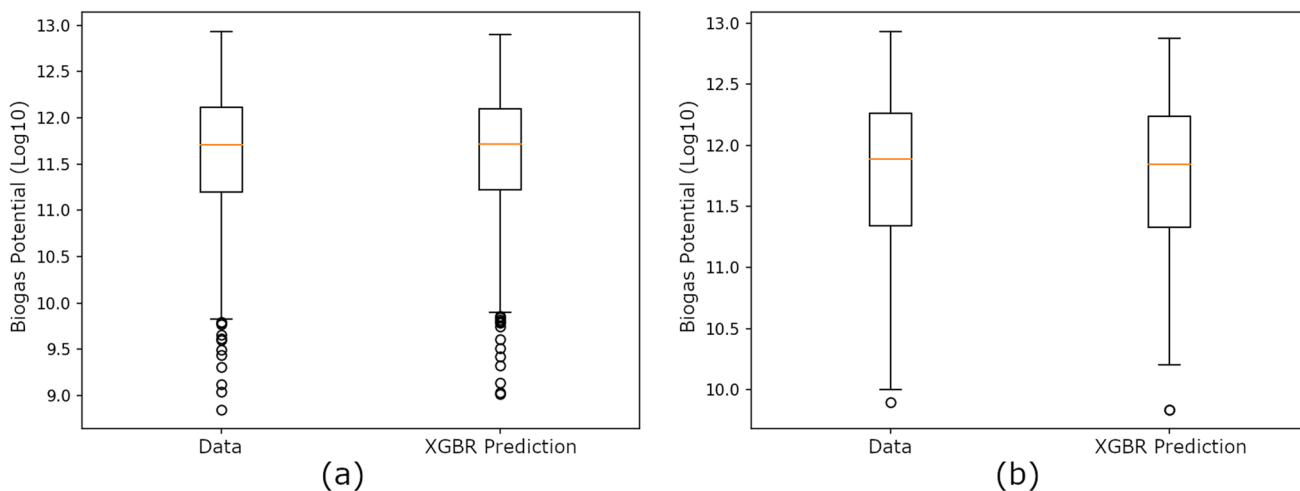
**Gradient boosting**

Gradient Boosting is a generic algorithm that sequentially assembles tree models. A generalization of the AdaBoost algorithm, gradient boosting allows any differentiable loss function. The difference between the predicted and actual values of the outcome variable is determined by fitting the tree to the loss function’s negative gradient. This allows it to optimize arbitrary differential loss functions (Friedman 2001). Gradient Boosting is an ensemble model that makes predictions by “boosting” the collection of subpar prediction models to create a more reliable model. The errors learned from previous base learners

are the focus of this model’s training of the current base learner (Othchere et al. 2022).

**Extreme gradient boosting**

XGBoost approach is one of the most popular gradient-boosted decision tree implementations and can solve the sparse data problem. The training process is accelerated by hardware acceleration and parallel processing in XGBoost, an optimized Gradient Boosting implementation. Regularization, weighted quantile sketches, parallel learning blocks, cache awareness, and out-of-core computing capabilities are all provided by XGBoost. L1/L2 penalties are used for regularization



**Fig. 4** Box plots of XGBR’s predictions with the theoretical calculations of animal-based biogas potential: **a** tenfold cross-validation results and **b** prediction results for 2020–2021

**Table 2** The performance scores for modeling animal-based CH<sub>4</sub> emissions for 2004–2021

Approach	Analysis	Algorithm	Train			Test/prediction		
			RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>
Tier1	Cross-validation (2004–2019)	Gradient Boosting	0.213 ± 0.003	0.151 ± 0.002	0.9898 ± 0.0003	0.238 ± 0.029	0.165 ± 0.017	0.9855 ± 0.0057
		XGBR	0.042 ± 0.003	0.022 ± 0.001	0.9996 ± 0.0000	0.186 ± 0.056	0.080 ± 0.017	<b>0.9914 ± 0.0043</b>
		AdaBoost	0.321 ± 0.002	0.254 ± 0.002	0.9768 ± 0.0009	0.329 ± 0.025	0.259 ± 0.020	0.9716 ± 0.0113
	Prediction (2020–2021)	Gradient Boosting	0.214	0.151	0.9897	0.209	0.153	0.6552
		XGBR	0.047	0.024	0.9995	0.046	0.034	<b>0.9835</b>
		AdaBoost	0.324	0.254	0.9764	0.333	0.254	0.1220
Tier2	Cross-validation (2004–2019)	Gradient Boosting	0.200 ± 0.004	0.159 ± 0.003	0.7872 ± 0.0068	0.217 ± 0.020	0.173 ± 0.011	0.7441 ± 0.0280
		XGBR	0.028 ± 0.002	0.018 ± 0.001	0.9959 ± 0.0004	0.119 ± 0.036	0.062 ± 0.007	<b>0.9191 ± 0.0437</b>
		AdaBoost	0.350 ± 0.006	0.289 ± 0.005	0.3530 ± 0.0190	0.354 ± 0.019	0.289 ± 0.019	0.3196 ± 0.0749
	Prediction (2020–2021)	Gradient Boosting	0.205	0.162	0.7766	0.211	0.170	0.7647
		XGBR	0.030	0.020	0.9953	0.066	0.055	<b>0.9773</b>
		AdaBoost	0.347	0.288	0.3612	0.365	0.297	0.2955

to control overfitting. Utilizing the weighted quantile sketch algorithm, it can also handle sparse data sets. The main idea behind the XGBoost algorithm is to divide features and add trees to grow a tree continuously. The predicted value of the sample is calculated by adding up the scores corresponding to each tree if a prediction is generated for each sample after training. This score is determined using the characteristics of this sample, which correspond to a leaf node in each tree (Chen and Guestrin 2016; Khan et al. 2023). XGBoost Regressor (XGBR) is used for regression problems.

The objective function of the XGBR algorithm is given in Eq. (4), while the regularization term in this function is given in Eq. (5).

$$OBJ = \sum_{i=1}^n L(\hat{y}_i, y_i) + \sum_{t=1}^k \Omega(f_t) \tag{4}$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \tag{5}$$

In Eqs. (4) and (5),  $L(\cdot)$  and  $\Omega(\cdot)$  refer to the loss function and the regularization term. The target value is  $y$ , and the predicted value is  $\hat{y}$ , the number of samples is  $n$ , and the current sample is  $i$  in the loss function.  $k$  denotes the number of trees in the current model,  $t$  denotes the current tree,  $T$  denotes the total number of leaf nodes, and  $\omega$  denotes the weight of each leaf. This term of regularization suppresses the complexity of the model that forms the objective function. Control parameters to prevent overfitting include  $\gamma$  and  $\lambda$ . The structure of XGBR is given in Fig. 2.

### Model evaluation

Testing the model on the test set after training should be conducted to determine its performance and generalizability. It is possible to use metrics for evaluating models in this context. An analysis of regression was conducted using the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination ( $R^2$ ), which are commonly used metrics for regression analysis. Equation (6)–(9) provides the equations for these metrics (Hajabdollahi Oudjerji et al. 2023).

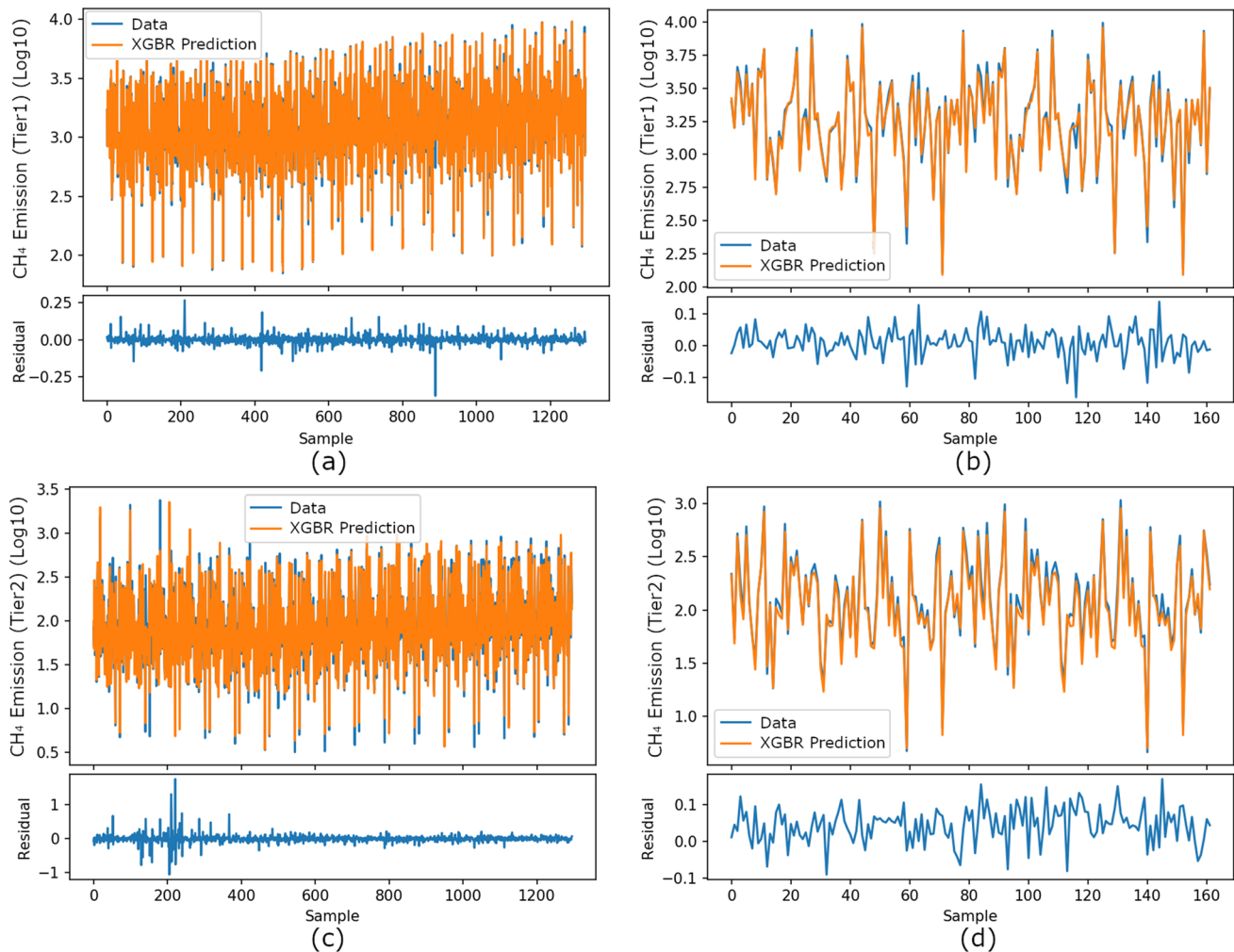
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{6}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{7}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{9}$$

According to Eqs. (6)–(9),  $y$  represents the target value,  $\hat{y}$  represents the predicted value,  $\bar{y}$  represents the mean of the target value, and  $n$  represents the sample size. MAPE is a relative measure based on percentage units instead of



**Fig. 5** Comparison of XGBR's predictions with the theoretical calculations of animal-based  $\text{CH}_4$  emissions: **a** tenfold cross-validation (tier1 approach), **b** prediction results for 2020–2021 (tier1 approach),

**c** tenfold cross-validation (tier2 approach), and **d** prediction results for 2020–2021 (tier2 approach)

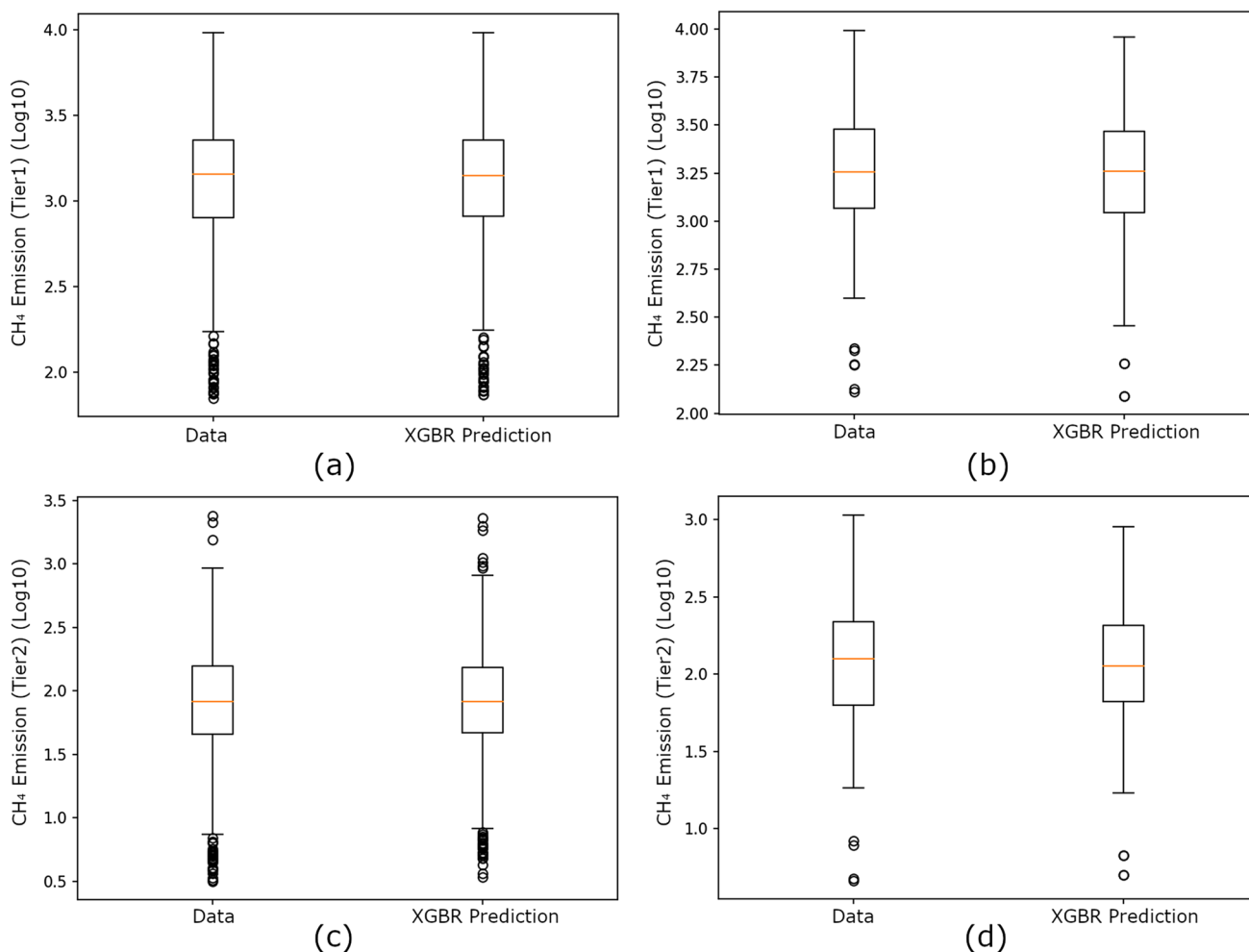
the variable's units to compare prediction accuracy between time-series models.

Cross-validation in ML evaluates a model's performance when it is applied to unseen data. Data is divided into multiple folds or subsets, and one fold serves as a validation set, while the other folds are used to train the model. The validation process is repeated several times, using different subsets of data each time. The model's performance on unseen data is then estimated by averaging the results. In this study, tenfold cross-validation was also used on the training data in addition to future predictions.

The statistical significance of relationships between ML algorithms' prediction values is assessed using statistical tests. The Wilcoxon rank-sum test confirms the inconsistencies between the model output and the actual value. If the  $p$  value of the model predictions is less than 0.05, the null hypothesis is rejected (Hayes et al. 2022).

## Boosting algorithm parameters

Default parameters in Sklearn and XGBoost libraries were used in all boosting algorithms. For Adaboost, the maximum number of predictors is 50, the learning rate is 1.0, and a linear loss function is used. For Gradient Boosting, the loss function is the square of the error, the learning rate is 0.1, the number of predictors is 100, the subsample rate is 1.0, the friedman\_mse function is used to measure the quality of a split, min\_samples\_split is 2, min\_samples\_leaf is 1, and the maximum depth of individual regression predictors is 3. For XGBR, the booster is gmtree, which uses tree-based models, 0.3 step size reduction used in updating to avoid overfitting, gamma is 0, max\_length is 6, min\_child\_weight is 1, 1.0 subsample ratio of training samples, sample\_method is uniform, tree generation method is faster histogram optimized approximate greedy algorithm.



**Fig. 6** Box plots of XGBR's predictions with the theoretical calculations of animal-based  $\text{CH}_4$  emissions: **a** tenfold cross-validation (tier1 approach), **b** prediction results for 2020–2021 (tier1 approach), **c**

tenfold cross-validation (tier2 approach), and **d** prediction results for 2020–2021 (tier2 approach)

## Results and discussion

This study used Adaboost, Gradient Boosting, and XGBR, popular boosting algorithms in the literature, to estimate the amount of biogas and  $\text{CH}_4$  emissions from animal sources. The proposed study includes two different analyses:

- The first one includes the biogas potential and  $\text{CH}_4$  emission values of each province in Turkey for the years 2004–2019 and the forecasts of each of these values for the years 2020–2021 with the boosting algorithms (scenario-1),
- The second one uses the same values for 2004–2021 and makes predictions for 2022–2024 (scenario-2).

Python programming language (ver. 3.10) with Scikit-learn library with boosting algorithms and XGBoost library was used for the execution of the algorithms.

In scenario-1, data for the 2004–2019 period were analyzed with a tenfold cross-validation analysis. Then, 1296 data for these years were used for training, and 162 data for the period 2020–2021 were tried to be predicted. The performance scores of cross-validation and predictions for animal-based biogas potential are shown in Table 1.

Table 1 shows that the XGBR algorithm is more successful in training and test scores in cross-validation and 2020–2021 predictions.

In comparing XGBR's predictions with the theoretical calculations of animal-based biogas potential, the tenfold cross-validation and prediction results for 2020–2021 are shown in Fig. 3a, b, respectively.

In Fig. 3a, the sample represents the record of each province for each year between 2004 and 2019, while in Fig. 3b, it represents each year between 2020 and 2021. In Fig. 3a, b, the value represents  $\log_{10}$  of the animal-based biogas potential of the relevant record. In Fig. 3, the theoretical

**Table 3** The MAPE scores for modeling animal-based biogas potential and CH<sub>4</sub> emissions for 2004–2021

Category	Analysis	Algorithm	MAPE (%)	
			Train	Test/prediction
Biogas potential	Cross-validation (2004–2019)	Gradient Boosting	1.84 ± 0.03	2.02 ± 0.12
		XGBR	0.22 ± 0.01	<b>0.63 ± 0.05</b>
		AdaBoost	3.70 ± 0.04	3.77 ± 0.11
	Prediction (2020–2021)	Gradient Boosting	1.80	1.69
		XGBR	0.23	<b>0.46</b>
		AdaBoost	3.68	3.55
CH <sub>4</sub> emissions (tier1)	Cross-validation (2004–2019)	Gradient Boosting	4.37 ± 0.05	4.73 ± 0.48
		XGBR	0.57 ± 0.03	<b>1.83 ± 0.24</b>
		AdaBoost	7.75 ± 0.06	7.88 ± 0.84
	Prediction (2020–2021)	Gradient Boosting	4.37	4.87
		XGBR	0.59	<b>1.07</b>
		AdaBoost	7.72	8.00
CH <sub>4</sub> emissions (tier2)	Cross-validation (2004–2019)	Gradient Boosting	9.12 ± 0.25	9.91 ± 1.11
		XGBR	1.02 ± 0.05	<b>3.52 ± 0.48</b>
		AdaBoost	16.71 ± 0.54	16.50 ± 1.78
	Prediction (2020–2021)	Gradient Boosting	9.30	8.51
		XGBR	1.08	<b>2.78</b>
		AdaBoost	16.56	14.48

**Table 4** Statistical comparisons of the XGBR algorithm predictions of biogas potential and CH<sub>4</sub> emissions with theoretical values

Category	Algorithm	The Wilcoxon rank-sum test			
		Cross-validation analysis		2020–2021 prediction	
		<i>h</i>	<i>p</i>	<i>h</i>	<i>p</i>
Biogas potential	Gradient Boosting	-	0.032	+	0.117
	XGBR	+	<b>0.894</b>	+	<b>0.521</b>
	AdaBoost	-	0.000	-	0.000
CH <sub>4</sub> emissions (tier1)	Gradient Boosting	+	0.700	+	0.165
	XGBR	+	<b>0.988</b>	+	<b>0.784</b>
	AdaBoost	-	0.023	-	0.000
CH <sub>4</sub> emissions (tier2)	Gradient Boosting	+	0.344	+	0.120
	XGBR	+	<b>0.937</b>	+	<b>0.304</b>
	AdaBoost	-	0.000	-	0.000

**Table 5** Theoretical values of biogas potential and CH<sub>4</sub> emissions of 5 major provinces in Turkey for the year 2021 and the predictions of the XGBR algorithm

Province	Theoretical values (Log10)			XGBR predictions (Log10)		
	Biogas potential	CH <sub>4</sub> emissions (tier1)	CH <sub>4</sub> emissions (tier2)	Biogas potential	CH <sub>4</sub> emissions (tier1)	CH <sub>4</sub> emissions (tier2)
İstanbul	11.278	3.069	1.787	11.209	3.078	1.824
Ankara	12.084	3.696	2.818	11.960	3.605	2.705
İzmir	12.300	3.933	2.747	12.323	3.919	2.742
Bursa	11.850	3.445	2.567	11.894	3.404	2.450
Antalya	12.869	3.298	2.125	12.825	3.296	2.057

calculation values of 1296 samples of biogas potential between 2004 and 2019 vary between 8.848 and 12.934. Residual refers to the difference between theoretical calculation and predicted values in statistical analysis. In the residual graphs given in Fig. 3, the values in 2004 fluctuate between -0.5 and 0.5, while the values for other years are monotonous. The residual graph of the 2020–2021 prediction shows that the values are monotonic and closer to 0.

As seen in Fig. 3, the XGBR predictions are consistent with the original curve, and the results can also be seen as a box plot in Fig. 4.

According to Fig. 4, the XGBR predictions (first quartile 11.20, median 11.71, and third quartile 12.05 for Fig. 4a, while first quartile 11.32, median 11.88, and third quartile 12.25 for Fig. 4b) overlap with the theoretical calculation values.

The performance scores of cross-validation and predictions for animal-based CH<sub>4</sub> emissions are shown in Table 2.

**Table 6** XGBR predictions for animal-based biogas potential and CH<sub>4</sub> emissions for 2024

Year	Province	XGBR predictions (Log10)		
		Biogas potential	CH <sub>4</sub> emissions (tier1)	CH <sub>4</sub> emissions (tier2)
2024	İstanbul	11.279	3.080	1.810
	Ankara	12.055	3.652	2.806
	İzmir	12.309	3.929	2.757
	Bursa	11.869	3.411	2.552
	Antalya	12.866	3.321	2.122

Table 2 shows that the XGBR algorithm gives the best results with  $R^2$  values of 0.9914 and 0.9191 in the cross-validation analysis for CH<sub>4</sub> emissions with tier1 and tier2 approaches, respectively. On the other hand, in terms of 2020–2021 predictions, it is seen that the XGBR algorithm gives the best results with  $R^2$  values of 0.9835 and 0.9773 in predicting CH<sub>4</sub> emissions with tier1 and tier2 approaches, respectively.

In comparing XGBR's predictions with the theoretical calculations of animal-based CH<sub>4</sub> emissions, the tenfold

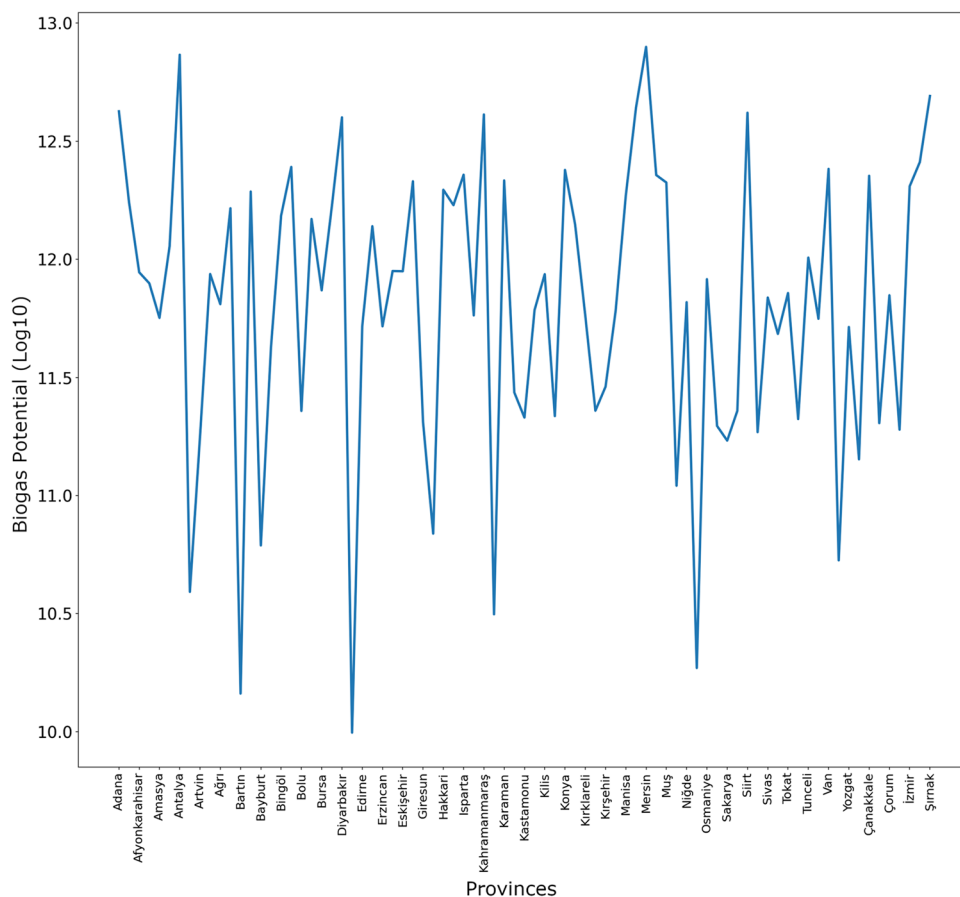
cross-validation and prediction results for 2020–2021 are shown in Fig. 5a–c and Fig. 5b–d, respectively.

In Fig. 5a–c, the sample represents the record of each province for each year between 2004 and 2019, while in Fig. 5b–d, it represents each year between 2020 and 2021. In Fig. 5, the value represents log<sub>10</sub> of the animal-based CH<sub>4</sub> emissions of the relevant record.

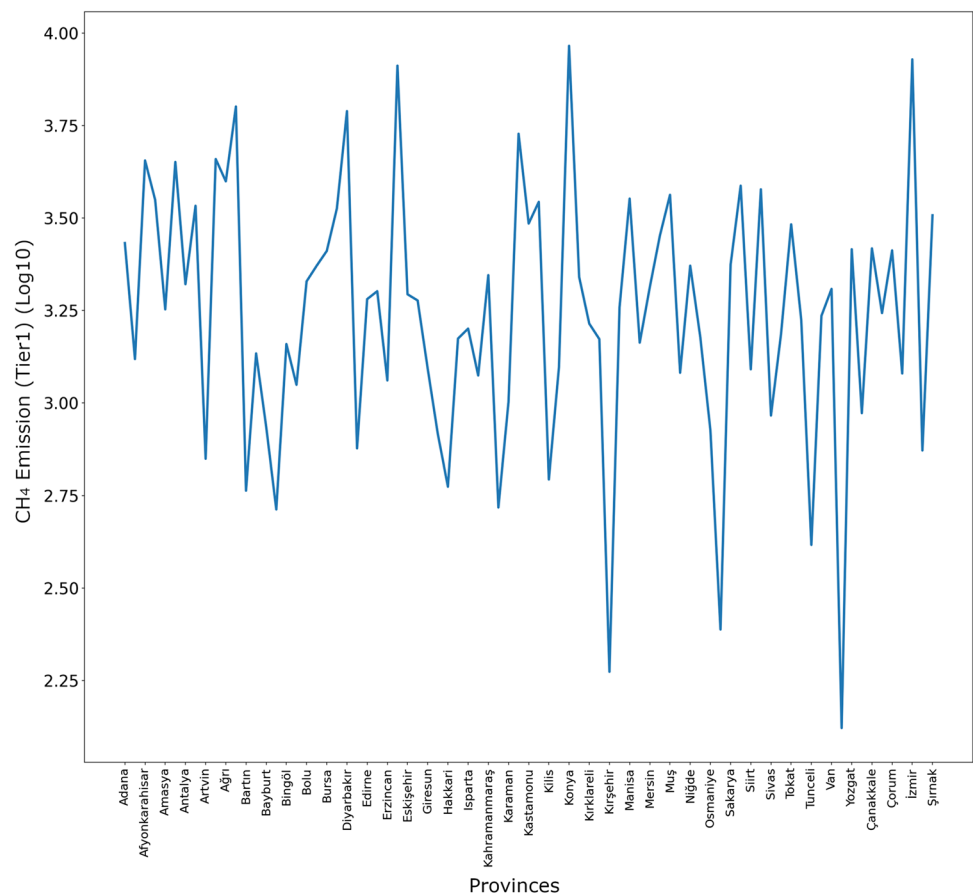
In Fig. 5a, the theoretical calculation values of CH<sub>4</sub> emissions (tier1 approach) for 16 years vary between 1.847 and 3.981. In the residual graphs given in Fig. 5a, b, the values for 2004–2019 fluctuate between –0.25 and 0.25, while the 2020–2021 prediction values vary between –0.1 and 0.1. In Fig. 5c, the theoretical calculation values of CH<sub>4</sub> emissions (tier2 approach) vary between 0.502 and 3.380. In the residual graphs given in Fig. 5c, d, the values for 2006 fluctuate between –1 and 1, while the values for other years monotonously approach zero. In the residual graph of the 2020–2021 prediction, the values vary between –0.1 and 0.1.

As seen in Fig. 5, the XGBR predictions are consistent with the original curve, and the results can also be seen as a box plot in Fig. 6.

In Fig. 6a, b, the XGBR predictions for CH<sub>4</sub> emissions by tier1 approach (first quartile 2.88, median 3.15, and third

**Fig. 7** XGBR predictions for animal-based biogas potential for 2024

**Fig. 8** XGBR predictions for animal-based CH<sub>4</sub> emissions (tier1 approach) for 2024



quartile 3.31 for Fig. 6a, while first quartile 3.05, median 3.26 and third quartile 3.47 for Fig. 6b) overlap with the theoretical calculation values. According to Fig. 6c, d, the XGBR predictions for CH<sub>4</sub> emissions by tier2 approach (first quartile 1.63, median 1.92, and third quartile 2.12 for Fig. 6c, while first quartile 1.80, median 2.10 and third quartile 2.33 for Fig. 6d) overlap with the theoretical calculation values.

Examining the error values as a percentage of the obtained models' results can enable a more comfortable evaluation of the success of the models. Therefore, the comparison of the predictions and theoretical values of the boosting algorithms for animal-based biogas potential and CH<sub>4</sub> emissions in terms of MAPE score are shown in Table 3.

As seen in Table 3, the error rates of XGBR predictions for biogas potential were relatively low, approximately 0.63% and 0.46% for cross-validation and future prediction, respectively. For CH<sub>4</sub> emission values, XGBR predictions are approximately 1.83% and 1.07% in tier1, while in tier2, they are approximately 3.52% and 2.78% for cross-validation and future prediction, respectively. Low percentage error values show that the XGBR algorithm is successful in modeling.

Table 4 contains statistical comparisons of the XGBR algorithm predictions of animal-derived biogas potential and CH<sub>4</sub> emissions with theoretical values.

The statistical test result is shown in Table 4 as *h* (“+,” accept; “–,” reject), and the test's *p*-value is the probability that the null hypothesis is true. When the Wilcoxon rank-sum test results in Table 4 are examined, it is seen that there is no significant difference between the results of this model and the theoretical values at the significance level of *p*=0.05.

Table 5 shows the theoretical values of biogas potential and CH<sub>4</sub> emissions of five major provinces in Turkey for the year 2021 and the predictions of the XGBR algorithm.

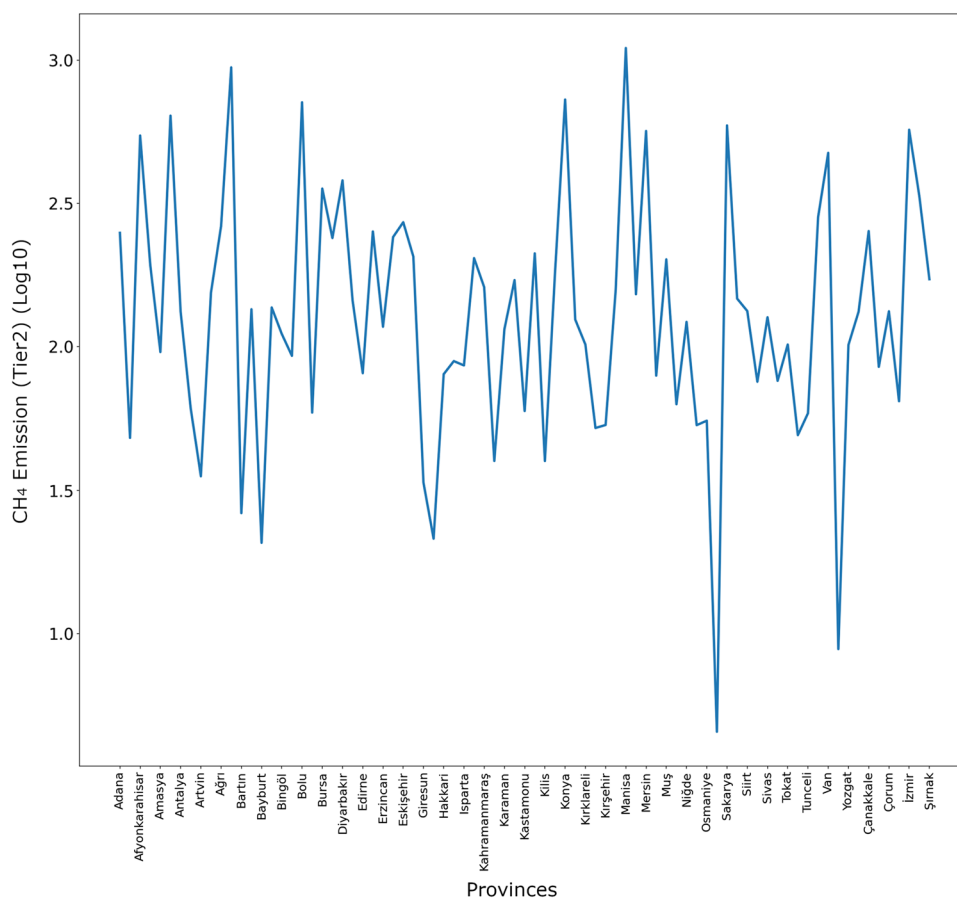
As seen in Table 5, the theoretical values of biogas potential and CH<sub>4</sub> emissions for 2021 for five major provinces in Turkey and the predictions of the XGBR algorithm are close to each other.

In scenario-2, the values for 2004–2021 are used for training, and predictions are made for 2022–2024.

Table 6 shows the XGBR predictions for animal-based biogas potential and CH<sub>4</sub> emissions of five significant provinces in Turkey for 2024.

According to the results in Table 6, the highest biogas potential among the five major provinces in Turkey in 2024

**Fig. 9** XGBR predictions for animal-based CH<sub>4</sub> emissions (tier2 approach) for 2024



belongs to Antalya, while the highest CH<sub>4</sub> emissions are estimated to be İzmir and Ankara for tier1 and tier2, respectively.

XGBR predictions for animal-based biogas potential for all provinces in Turkey in 2024 are shown in Fig. 7.

As seen in Fig. 7, the highest value in the 2024 animal-based biogas potential prediction for all provinces of Turkey belongs to Mersin province, while the lowest value belongs to Düzce province.

XGBR predictions for animal-based CH<sub>4</sub> emissions (tier1 approach) for all provinces in Turkey in 2024 are shown in Fig. 8.

As seen in Fig. 8, the highest value in the animal-based CH<sub>4</sub> emission (tier1 approach) prediction for 2024 in all provinces of Turkey belongs to Konya province, while the lowest value belongs to Yalova province.

XGBR predictions for animal-based CH<sub>4</sub> emissions (tier2 approach) for all provinces in Turkey in 2024 are shown in Fig. 9.

As seen in Fig. 9, the highest value in animal-based CH<sub>4</sub> emission (tier2 approach) estimation for 2024 in all provinces of Turkey belongs to Manisa, while the lowest value belongs to Rize.

The greenhouse gas inventory for Turkey is calculated as 564.4 Mt CO<sub>2e</sub> in 2021, increasing by 7.7% compared to 2020. Total greenhouse gas emissions per person increased by 0.4 tons compared to the previous year, reaching 6.7 tons of CO<sub>2</sub> in 2021 (TUIK 2022). In this regard, it is thought that studies such as the current study will support Turkey's 2050 climate change strategy and 2030 action plan preparations and the national contribution declaration, the Climate Change Directorate of the Ministry of Environment, Urbanization and Climate Change, and the relevant institutions within the scope of the United Nations Development Programme.

Although the study contains detailed information and analysis regarding animal husbandry in Turkey, it can only be considered limited to this area. However, it can also be improved in this context using different algorithms, parameters, etc. (e.g., agricultural wastes, geographical location, climatic conditions). The data in the study was obtained from the country's statistical institution. These data-induced errors can also affect the applicability of the model. The information obtained as a result of the study will be statistically valuable.



## Conclusion

Emissions from the livestock sector have an essential place in climate change. Sustainable manure management and biogas production are crucial for countries to solve this problem. This study used boosting algorithms to investigate the animal-based biogas potential and CH<sub>4</sub> emissions using tier1 and tier2 approaches in all Turkey provinces from 2004 to 2021. The XGBR algorithm was the most successful in predicting animal-based biogas potential and CH<sub>4</sub> emissions, with MAPE ranging from 0.46 to 2.78%. The study also predicted the biogas potential of five major cities in Turkey for 2022–2024. The European Union aims to be a global role model in combating the climate crisis and achieving sustainable development goals. In this regard, countries want to reach climate neutral by 2050 with the Green Deal agreement, and it is thought that the prediction model proposed in this study can guide researchers for the coming years.

**Author contribution** All authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by Ihsan Pence, Kazım Kumaş, and Melike Siseci Cesmeli. The first draft of the manuscript was written by Ihsan Pence and Kazım Kumaş. Visualizations were performed by Melike Siseci Cesmeli. Review and editing were carried out by Ali Özhan Akyüz. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

**Data availability** All data generated or analyzed during this study are included in this published article.

## Declarations

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Almomani F (2020) Prediction of biogas production from chemically treated co-digested agricultural waste using artificial neural network. *Fuel* 280:118573. <https://doi.org/10.1016/J.FUEL.2020.118573>
- Andrade Cruz I, Chuenchart W, Long F, Surendra KC, Renata Santos Andrade L, Bilal M, Liu H, Tavares Figueiredo R, Khanal SK, Fernando Romanholo Ferreira L (2022) Application of machine learning in anaerobic digestion: perspectives and challenges. *Bioresour Technol* 345:126433. <https://doi.org/10.1016/J.BIORTECH.2021.126433>
- Avcioglu AO, Türker U (2012) Status and potential of biogas energy from animal wastes in Turkey. *Renew Sustain Energy Rev* 16:1557–1561. <https://doi.org/10.1016/J.RSER.2011.11.006>
- Baek CY, Lee KM, Park KH (2014) Quantification and control of the greenhouse gas emissions from a dairy cow system. *J Clean Prod* 70:50–60. <https://doi.org/10.1016/j.jclepro.2014.02.010>
- Basak JK, Arulmozhi E, Moon BE, Bhujel A, Kim HT (2022) Modelling methane emissions from pig manure using statistical and machine learning methods. *Air Qual Atmos Health* 1–15. <https://doi.org/10.1007/s11869-022-01169-0>
- Can A (2020) The statistical modeling of potential biogas production capacity from solid waste disposal sites in Turkey. *J Clean Prod* 243:118501. <https://doi.org/10.1016/j.jclepro.2019.118501>
- Ceylan AB, Aydın L, Nil M, Mamur H, Polatoğlu İ, Sözen H (2023) A new hybrid approach in selection of optimum establishment location of the biogas energy production plant. *Biomass Convers Biorefin* 13:5771–5786. <https://doi.org/10.1007/S13399-021-01532-8>
- Chandra Manna M, Rahman MM, Naidu R, Sahu A, Bhattacharjya S, Wanjarı RH, Patra AK, Chaudhari SK, Majumdar K, Khanna SS (2018) Bio-waste management in subtropical soils of India: future challenges and opportunities in agriculture. *Adv Agron* 152:87–148. <https://doi.org/10.1016/BS.AGRON.2018.07.002>
- Chen Z, An C, Fang H, Zhang Y, Zhou Z, Zhou Y, Zhao S (2020) Assessment of regional greenhouse gas emission from beef cattle production: a case study of Saskatchewan in Canada. *J Environ Manag* 264:110443. <https://doi.org/10.1016/j.jenvman.2020.110443>
- Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Assoc Comput Mach, pp 785–794. <https://doi.org/10.1145/2939672.2939785>
- Cheng D, Liu Y, Shehata E, Feng Y, Lin H, Xue J, Li Z (2021) In-feed antibiotic use changed the behaviors of oxytetracycline, sulfamerazine, and ciprofloxacin and related antibiotic resistance genes during swine manure composting. *J Hazard Mater* 402:123710. <https://doi.org/10.1016/J.JHAZMAT.2020.123710>
- Chiu MC, Wen CY, Hsu HW, Wang WC (2022) Key wastes selection and prediction improvement for biogas production through hybrid machine learning methods. *Sustain Energy Technol Assess* 52:102223. <https://doi.org/10.1016/J.SETA.2022.102223>
- Das NG, Sarker NR, Haque MN (2020) An estimation of greenhouse gas emission from livestock in Bangladesh. *J Adv Vet Anim Res* 7:133. <https://doi.org/10.5455/JAVAR.2020.G402>
- De Clercq D, Jalota D, Shang R, Ni K, Zhang Z, Khan A, Wen Z, Caicedo L, Yuan K (2019) Machine learning powered software for accurate prediction of biogas production: a case study on industrial-scale Chinese production data. *J Clean Prod* 218:390–399. <https://doi.org/10.1016/J.JCLEPRO.2019.01.031>
- Dong H, Mangino J, McAllister TA, Hatfeld JL, Johnson DE, Lasey KR, de Lima MA, Romanovskaya A (2006) Emissions from livestock and manure management. In: Guidelines for National greenhouse gas inventories, Intergovernmental Panel on Climate Change (IPCC 2006). <https://www.ipcc-nggip.iges.or.jp/public/>

- 2006gl/pdf/4\_Volume4/V4\_10\_Ch10\_Livestock.pdf. Accessed 12 Nov 2022
- Elmaz F, Yücel Ö, Mutlu AY (2020) Predictive modeling of biomass gasification with machine learning-based regression methods. *Energy* 191:116541. <https://doi.org/10.1016/J.ENERGY.2019.116541>
- Erdin C, Ozkaya G (2019) Turkey's 2023 energy strategies and investment opportunities for renewable energy sources: site selection based on ELECTRE. *Sustainability* 11:2136. <https://doi.org/10.3390/su11072136>
- Erdogdu AE, Polat R, Ozbay G (2019) Pyrolysis of goat manure to produce bio-oil. *Eng Sci Technol Int J* 22:452–457. <https://doi.org/10.1016/J.JESTCH.2018.11.002>
- Fajobi MO, Lasode OA, Adeleke AA, Ikubanni PP, Balogun AO (2022) Effect of biomass co-digestion and application of artificial intelligence in biogas production: a review. *Energy Sources Part A* 44:5314–5339. <https://doi.org/10.1080/15567036.2022.2085823>
- Font-Palma C (2019) Methods for the treatment of cattle manure—a review. *C (basel)* 5:27. <https://doi.org/10.3390/c5020027>
- Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55:119–139. <https://doi.org/10.1006/JCSS.1997.1504>
- Friedman JH (2001) Greedy function approximation : a gradient boosting machine. *Ann Stat* 29:1189–1232
- Ganaie MA, Hu M, Malik AK, Tanveer M, Suganthan PN (2022) Ensemble deep learning: a review. *Eng Appl Artif Intell* 115:105151. <https://doi.org/10.1016/j.engappai.2022.105151>
- Géron A (2019) Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow : concepts, tools, and techniques to build intelligent systems. O'Reilly Media, California
- Gonçalves Neto J, Vidal Ozorio L, Campos de Abreu TC, Ferreira dos Santos B, Pradelle F (2021) Modeling of biogas production from food, fruits and vegetables wastes using artificial neural network (ANN). *Fuel* 285:119081. <https://doi.org/10.1016/J.FUEL.2020.119081>
- Gündoğan B, Koçar G (2022) Potential usability of *Cynara cardunculus* L. residues in biogas production in various regions of Turkey. *Bioenergy Res* 15:1894–1907. <https://doi.org/10.1007/s12155-021-10375-0>
- Hajabdollahi Ouderji Z, Gupta R, Mckeown A, Yu Z, Smith C, Sloan W, You S (2023) Integration of anaerobic digestion with heat pump: machine learning-based technical and environmental assessment. *Bioresour Technol* 369:128485. <https://doi.org/10.1016/J.BIORTECH.2022.128485>
- Hayes A, May Murray L, Catherine Stanton I, Zhang L, Snape J, Hugo Gaze W, Kaye Murray A (2022) Predicting selection for antimicrobial resistance in UK wastewater and aquatic environments: ciprofloxacin poses a significant risk. *Environ Int* 169:107488. <https://doi.org/10.1016/j.envint.2022.107488>
- Herrera AMN, Esteves EMM, Morgado CRV, Esteves VPP (2021) Carbon footprint analysis of bioenergy production from cattle manure in the Brazilian central-west. *Bioenergy Res* 14:1265–1276. <https://doi.org/10.1007/s12155-020-10216-6>
- Heydari B, Abdollahzadeh Sharghi E, Rafiee S, Mohtasebi SS (2021) Use of artificial neural network and adaptive neuro-fuzzy inference system for prediction of biogas production from spearmint essential oil wastewater treatment in up-flow anaerobic sludge blanket reactor. *Fuel* 306:121734. <https://doi.org/10.1016/J.FUEL.2021.121734>
- Heydari A, Alborzi ZS, Amini Y, Hassanvand A (2023) Configuration optimization of a renewable hybrid system including biogas generator, photovoltaic panel and wind turbine: particle swarm optimization and genetic algorithms. *Int J Mod Phys C* 34:2350069. <https://doi.org/10.1142/S0129183123500699>
- Hörtenhuber SJ, Seiringer M, Theurl MC, Größbacher V, Piringer G, Kral I, Zollitsch WJ (2022) Implementing an appropriate metric for the assessment of greenhouse gas emissions from livestock production: a national case study. *Animal* 16:100638. <https://doi.org/10.1016/J.ANIMAL.2022.100638>
- Huo L, Zhao L, Yao Z, Luo J, Zhang P, Xie T, Jia J, Deng Y, Wei X (2021) Potentiality of agricultural biomass energy for greenhouse gas emission reduction. *Trans Chin Soc Agric Mach* 37:179–187. <https://doi.org/10.11975/j.issn.1002-6819.2021.22.020>
- Ibidhi R, Kim T-H, Bharanidharan R, Lee H-J, Lee Y-K, Kim N-Y, Kim K-H (2021) Developing country-specific methane emission factors and carbon fluxes from enteric fermentation in South Korean dairy cattle production. *Sustainability* 13:9133. <https://doi.org/10.3390/su13169133>
- Jeong K, Abbas A, Shin J, Son M, Kim YM, Cho KH (2021) Prediction of biogas production in anaerobic co-digestion of organic wastes using deep learning models. *Water Res* 205:117697. <https://doi.org/10.1016/j.watres.2021.117697>
- Jones D, Moore C (2023) Renewables beat fossil fuels: a half-yearly analysis of Europe's electricity transition. Available at: <https://ember-climate.org/insights/research/eu-electricity-analysis-h1-2020/>. Accessed June 10, 2023
- Karaaslan A, Gezen M (2022) The evaluation of renewable energy resources in Turkey by integer multi-objective selection problem with interval coefficient. *Renew Energy* 182:842–854. <https://doi.org/10.1016/J.RENENE.2021.10.053>
- Khan MU, Ahmad M, Sultan M, Sohoo I, Ghimire PC, Zahid A, Sarwar A, Farooq M, Sajjad U, Abdeslahian P, Yousaf M (2021) Biogas production potential from livestock manure in Pakistan. *Sustainability (switzerland)* 13:6751. <https://doi.org/10.3390/su13126751>
- Khan M, Raza Naqvi S, Ullah Z, Ali Ammar Taqvi S, Nouman Aslam Khan M, Farooq W, Taqi Mehran M, Juchelková D, Štěpanec L (2023) Applications of machine learning in thermochemical conversion of biomass—a review. *Fuel* 332:126055. <https://doi.org/10.1016/j.fuel.2022.126055>
- Khoshgoftar Manesh MH, Rezazadeh A, Kabiri S (2020) A feasibility study on the potential, economic, and environmental advantages of biogas production from poultry manure in Iran. *Renew Energy* 159:87–106. <https://doi.org/10.1016/J.RENENE.2020.05.173>
- Kim D, Kim K-T, Park Y-K (2020) A comparative study on the reduction effect in greenhouse gas emissions between the combined heat and power plant and boiler. *Sustainability* 12:5144. <https://doi.org/10.3390/su12125144>
- Liu T, Ferrari G, Pezzuolo A, Alengebawy A, Jin K, Yang G, Li Q, Ai P (2023) Evaluation and analysis of biogas potential from agricultural waste in Hubei Province, China. *Agric Syst* 205:103577. <https://doi.org/10.1016/j.agsy.2022.103577>
- Ludlow J, Jalil-Vega F, Schmidt Rivera X, Garrido RA, Hawkes A, Staffell I, Balcombe P (2021) Organic waste to energy: resource potential and barriers to uptake in Chile. *Sustain Prod Consum* 28:1522–1537. <https://doi.org/10.1016/j.spc.2021.08.017>
- Maroušek J, Strunecký O, Kolář L, Vochozka M, Kopecký M, Maroušková A, Batt J, Poliak M, Šoch M, Bartoš P, Klieštík T, Filip M, Konvalina P, Moudrý J, Peterka J, Suchý K, Zoubek T, Cera E (2020) Advances in nutrient management make it possible to accelerate biogas production and thus improve the economy of food waste processing. *Energy Source Part A* 1–10. <https://doi.org/10.1080/15567036.2020.1776796>
- Melikoglu M, Menekse ZK (2020) Forecasting Turkey's cattle and sheep manure based biomethane potentials till 2026. *Biomass Bioenergy* 132:105440. <https://doi.org/10.1016/J.BIOMBIOE.2019.105440>
- Najafi B, Faizollahzadeh Ardabili S (2018) Application of ANFIS, ANN, and logistic methods in estimating biogas production from spent mushroom compost (SMC). *Resour Conserv Recycl* 133:169–178. <https://doi.org/10.1016/J.RESCONREC.2018.02.025>

- Nehra M, Jain S (2023) Estimation of renewable biogas energy potential from livestock manure: a case study of India. *Bioresour Technol Rep* 22:101432. <https://doi.org/10.1016/j.biteb.2023.101432>
- Ngwabie NM, Chungong BN, Yengong FL (2018) Characterisation of pig manure for methane emission modelling in Sub-Saharan Africa. *Biosyst Eng* 170:31–38. <https://doi.org/10.1016/j.biosysteng.2018.03.009>
- Ocak S, Acar S (2021) Biofuels from wastes in Marmara Region, Turkey: potentials and constraints. *Environ Sci Pollut Res* 28:66026–66042. <https://doi.org/10.1007/s11356-021-15464-3>
- OECD (2015) Climate Change Risks and Adaptation. OECD. <https://doi.org/10.1787/9789264234611-en>
- Otchere DA, Ganat TOA, Ojoro JO, Tackie-Otoo BN, Taki MY (2022) Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions. *J Pet Sci Eng* 208:109244. <https://doi.org/10.1016/j.petrol.2021.109244>
- Pence I, Kumaş K, Siseçi MC, Akyüz A (2022) Modeling of energy and emissions from animal manure using machine learning methods: the case of the Western Mediterranean Region, Turkey. *Environ Sci Pollut Res* 30:22631–22652. <https://doi.org/10.1007/s11356-022-23780-5>
- Riño B, García-González MC (2015) Greenhouse gas emissions of an on-farm swine manure treatment plant - comparison with conventional storage in anaerobic tanks. *J Clean Prod* 103:542–548. <https://doi.org/10.1016/j.jclepro.2014.07.007>
- Senocak AA, Guner Goren H (2022) Forecasting the biomass-based energy potential using artificial intelligence and geographic information systems: a case study. *Eng Sci Technol Int J* 26:100992. <https://doi.org/10.1016/J.JESTCH.2021.04.011>
- Senocak AA, Guner Goren H (2023) Three-phase artificial intelligence-geographic information systems-based biomass network design approach: a case study in Denizli. *Appl Energy* 343:121214. <https://doi.org/10.1016/j.apenergy.2023.121214>
- Şenol H, Ali Dereli M, Özbilgin F (2021) Investigation of the distribution of bovine manure-based biomethane potential using an artificial neural network in Turkey to 2030. *Renew Sust Energy Rev* 149:111338. <https://doi.org/10.1016/J.RSER.2021.111338>
- Sharafi S, Kazemi A, Amiri Z (2023) Estimating energy consumption and GHG emissions in crop production: A machine learning approach. *J Clean Prod* 408:137242. <https://doi.org/10.1016/J.JCLEPRO.2023.137242>
- Shin J, Hong SG, Kim SC, Yang JE, Lee SR, Li FZ (2016) Estimation of potential methane production through the mass balance equations from agricultural biomass in Korea. *Appl Biol Chem* 59:765–773. <https://doi.org/10.1007/s13765-016-0224-1>
- Stolarski MJ, Warmiński K, Krzyżaniak M, Olba-Zięty E, Akinca M (2020) Bioenergy technologies and biomass potential vary in Northern European countries. *Renew Sust Energy Rev* 133:110238. <https://doi.org/10.1016/J.RSER.2020.110238>
- Sun H, Wang E, Li X, Cui X, Guo J, Dong R (2021) Potential biomethane production from crop residues in China: contributions to carbon neutrality. *Renew Sust Energy Rev* 148:111360. <https://doi.org/10.1016/J.RSER.2021.111360>
- Tongwane MI, Moeletsi ME (2021) Provincial cattle carbon emissions from enteric fermentation and manure management in South Africa. *Environ Res* 195:110833. <https://doi.org/10.1016/J.ENVRES.2021.110833>
- Tufaner F, Demirci Y (2020) Prediction of biogas production rate from anaerobic hybrid reactor by artificial neural network and nonlinear regressions models. *Clean Technol Environ Policy* 22:713–724. <https://doi.org/10.1007/s10098-020-01816-z>
- TUIK (2022) Turkish Statistical Institute. Available at: <https://data.tuik.gov.tr/Kategori/GetKategori?p=Nufus-ve-Demografi-109>. Accessed January 5, 2022
- Usack JG, Van Doren LG, Posmanik R, Tester JW, Angenent LT (2019) Harnessing anaerobic digestion for combined cooling, heat, and power on dairy farms: an environmental life cycle and techno-economic assessment of added cooling pathways. *J Dairy Sci* 102:3630–3645. <https://doi.org/10.3168/JDS.2018-15518>
- Wang Y, Wehrle L, Banerjee A, Shi Y, Deutschmann O (2021) Analysis of a biogas-fed SOFC CHP system based on multi-scale hierarchical modeling. *Renew Energy* 163:78–87. <https://doi.org/10.1016/J.RENENE.2020.08.091>
- Yurtkuran S (2021) The effect of agriculture, renewable energy production, and globalization on CO2 emissions in Turkey: a bootstrap ARDL approach. *Renew Energy* 171:1236–1245. <https://doi.org/10.1016/J.RENENE.2021.03.009>
- Zaidi AA, RuiZhe F, Shi Y, Khan SZ, Mushtaq K (2018) Nanoparticles augmentation on biogas yield from microalgal biomass anaerobic digestion. *Int J Hydrogen Energy* 43:14202–14213. <https://doi.org/10.1016/J.IJHYDENE.2018.05.132>
- Zhang W, Wu Y, Calautit JK (2022) A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment. *Renew Sust Energy Rev* 167:112704. <https://doi.org/10.1016/j.rser.2022.112704>
- Zhang C, Nie J, Yan X (2023) Estimation of biomass utilization potential in China and the impact on carbon peaking. *Environ Sci Pollut Res* 30:94255–94275. <https://doi.org/10.1007/S11356-023-28891-1>
- Zubir MA, Bong CPC, Ishak SA, Ho WS, Hashim H (2022) The trends and projections of greenhouse gas emission by the livestock sector in Malaysia. *Clean Technol Environ Policy* 24:363–377. <https://doi.org/10.1007/s10098-021-02156-2>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.