scientific reports

Check for updates

OPEN

Optimizing the early-stage of composting process emissions – artificial intelligence primary tests

Joanna Rosik[1,2,4], Maciej Karczewski[3] & Sylwia Stegenta-Dąbrowska[2,4]✉

Although composting has many advantages in treating organic waste, many problems and challenges are still associated with emissions, like NH_3, CO and H_2S, as well as greenhouse gases such as CO_2. One promising approach to enhancing composting conditions is using novel analytical methods based on artificial intelligence. To predict and optimize the emissions (CO, CO_2, H_2S, NH_3) during the early-stage of composting process machine learning (ML) models were utilized. Data about emissions from laboratory composting with compost's biochar with different incubation (50, 60, 70 °C) and biochar doses (0, 3, 6, 9, 12, 15% dry mass) were used for ML models selections and training. ML models such as acritical neural network (ANN, Bayesian Regularized Neural Network; R^2 accuracy CO:0.71, CO_2:0.81, NH_3:0.95, H_2S:0.72) and decision tree (DT, RPART; R^2 accuracy CO:0.69, CO_2:0.80, NH_3:0.93, H_2S:0.65) have demonstrated satisfactory results. The ML models to predict CO and H_2S during composting were demonstrated for the first time. Utilizing emission data to predict other noxious gases presents a cost-effective and expeditious alternative to the empirical analysis of compost properties.

Keywords Machine learning 1, Biochar application 2, Greenhouse gases 3, Composting optimizing 4

Composting process is one of the most popular ways to manage biodegradable waste because it is highly effective, low-risk, and environmentally beneficial. Despite these advantages, composting process may cause emissions of hazardous odors and gases like NH_3, H_2S, CO and CO_2 which is especially environmentally disadvantageous[1–3]. Furthermore, these gases pose a risk to the employees working in the composting plant. The monitoring of gas emissions, such as NH_3, H_2S, and CO, is often disregarded within the composting industry. However, existing data suggests that composting's greenhouse gas (GHG) emissions are substantial, equating to 183 kg CO2e per ton of waste[4]. Furthermore, these emissions could harm employees when not monitored and predicted accurately. High concentrations e.g. CO[5] in poorly ventilated areas could adversely affect people's health and even death. Hence, it is imperative to ascertain the optimal composting process conditions for minimizing gaseous emissions, which can be achieved throught the adjustment of process parameters or by employing mathematical simulation. Currently, a popular solution used to reduce emissions of GHG is biochar, which can retain gaseous substances on its surface due to its physicochemical properties[6–8]. There is still a lack of research to determine the ideal parameters for biochar production, dosage and incubation temperature of the composted material. In addition, the relationships between these parameters are very complex and depends on human oversight, which complicates compost quality assessment and the composting process's management. Therefore, the development of automated systems for controlling the emissions through composting with biochar is essential to enhance uniformity and effectiveness of the process.

Artificial intelligence (AI), such as machine learning (ML), is becoming increasingly common in optimizing multiple processes. With ML, it is possible to assess and improve response conditions and maximize operational efficiency by optimizing necessary parameters, especially in agricultural and environmental sciences[9,10]. ML methods widely used in waste management include models such

[1]Institute of Environmental Engineering, The Faculty of Environmental Engineering and Geodesy, Wrocław University of Environmental and Life Sciences, Grunwaldzki Square 24, Wrocław 50-363, Poland. [2]Department of Applied Bioeconomy, Wrocław University of Environmental and Life Sciences, 37a Chełmońskiego Str, Wrocław 51-630, Poland. [3]Department of Applied Mathematics, Wrocław University of Environmental and Life Sciences, 53 Grunwaldzki Sq., 50-363 Wrocław, Poland. [4]Joanna Rosik and Sylwia Stegenta-Dąbrowska have contributed equally to this work. ✉email: sylwia.stegenta-dabrowska@upwr.edu.pl

as artificial neural networks (ANN), support vector machines (SVM), decision trees (DT), K Nearest Neighbor (kNN), radial basis function (RBF) and various other ensemble learning techniques[9,10].

ML was primarily used in the composting process to predict the quality[11] and quantity[12] of the resulting compost, as well as to monitor the process itself, e.g. controlling heat loss[13]. Ding et al. examined the possibilities of using ML models to optimize the maturity of kitchen waste composting. The study revealed that different stages of the composting process should be modeled using various parameters and the model-based system exhibited better maturity of the final material[14]. Despite the increasing use of machine learning in composting, there is a relative scarcity of research on adding biochar, particularly in relation to the prediction of greenhouse gas emissions. Predicting compost gas emissions, especially with biochar application, is more challenging. Li et al. developed Random Forest Algorithm to predict CO_2[15]. Artificial Neural Network (ANN) was utilized for predicting NH_3 emissions during composting sewage sludge with straw[13]. However, the presented models necessitate costly and labor-intensive tests to assess substrate properties such as C/N, N-NH_4, lignin content, and hemicellulose—the parameters facilitate emission prediction. In contrast, gas sensor-based GHG predictions presented in this study offer the advantage of complete automation and, most significantly, immediate online results from algorithms.

The present study aims to compare different machine learning (ML) models to predict the emissions (CO, CO_2, H_2S, NH_3) during the first 10 days of composting with compost's biochar addition. Data about everyday emissions for modeling were collected during laboratory composting with compost's biochar with different incubation temperature. The use of ML for optimizations and limitations of the emissions during early-stage composting has good potential and can be used to improve the safety of the process. Utilizing emission data to predict other noxious gases presents a cost-effective and expeditious alternative to the empirical analysis of compost properties.

Results

Prediction of the gaseous emissions during composting with composts' biochar (machine learning)

Ten kinds of classifiers, Linear Regression, Generalized Linear model, Random Forest, SVM with Linear Kernel, SVM with Radial Basis Function Kernel, k-Nearest Neighbours, Bayesian Regularized Neural Network, RPART, Generalized Boosted Regression Models and Extreme Gradient Boosting Tree were trained using collected data to evaluate the practicality of the classification model in predicting gaseous emissions output. Determination coefficients R^2 and RMSE were used to determine the model's effectiveness; the results are shown in Table 1. The best results ($R^2 \geq 0.6$) for each emission were observed for the Bayesian Regularized Neural Network. A comparable good performance was also characteristic of RPART. These models were also characterized by a low RMSE (CO < 380; CO_2 < 120; H_2S < 40 NH_3 < 80), while its values are dependent on the measured emission value of the gas, hence the significant discrepancies between the observed results. In addition, the best accuracy was observed in NH_3 emission, where $R_2 > 0.9$. This demonstrates not only the good fit of the model to the results obtained during the tests but also the high potential for predicting emissions of this gas from the remaining input data. A high potential for predicting NH_3 has also been observed in the literature. Xie et al. used models based on artificial neural networks, the Adaptive Neuro Fuzzy Inference System (ANFIS), to predict ammonia emissions from pig-fattening houses using various inputs. He contrasted the results with models, such as the Multiple Linear Regression Model and Backpropagation. With ANFIS, it was possible to obtain high R^2 values (> 0.6) during both summertime and wintertime[16]. Küçüktopcu et al. used ANFIS and Multilayer Perception (MLP) models to model NH_3 emissions on poultry farms. Modeling was performed using input data such as indoor air temperature, air humidity, air flow, NH_3 emission concentrations, litter moisture, litter pH and litter surface temperatures. Input data were used for modeling in different configurations, while the best results were obtained for the ANFIS model with subtractive clustering ($R^2 = 0.910$; RMSE = 0.919) in the input data configuration using litter moisture, air temperature and airflow[17]. Models also show high potential for CO_2 prediction. Li et al. used the AdaBoost, Bagging, Gradient Boost, Random Forest, k-Nearest Neighbors and Decision Tree models. The k-Nearest Neighbors model achieved the highest prediction accuracy, with an RMSE of 54.9. However, the

	CO		CO_2		NH_3		$H2_S$	
Model	R^2	**RMSE**	R^2	**RMSE**	R^2	**RMSE**	R^2	**RMSE**
Linear regression	0.304	376.870	0.538	120.130	0.350	36.010	0.141	83.533
Random forest	0.463	331.256	0.741	89.841	0.918	12.791	0.567	59.277
SVM with linear Kernel	0.255	389.928	0.503	124.443	0.212	39.644	0.072	86.811
SVM with RBF Kernel	0.636	272.579	0.776	83.699	0.900	14.125	0.602	56.888
k-nearest neighbors	0.466	330.187	0.730	91.852	0.895	14.453	0.261	77.461
Bayesian regularized neural network	0.710	243.318	0.808	77.465	0.948	10.159	0.715	48.111
RPART	0.693	250.324	0.802	78.562	0.930	11.796	0.648	53.459
Generalized boosted regression models	0.595	287.527	0.764	79.493	0.899	14.163	0.584	58.104
Extreme gradient boosting tree	0.309	375.754	0.798	85.764	0.793	20.326	0.486	64.608
Partial least squares regression	--	–	0.544	119.348	0.360	35.737	0.149	83.131

Table 1. Comparisons between particular models by values of R squared and RMSE.

authors have pointed out that the regression model's prediction granularity is too sensitive to changes in data distribution, resulting in less-than-ideal prediction performance[15]It needs to be underlined that the use of ML models to predict CO and H_2S during composting was demonstrated for the first time with sufficient accuracy using a Bayesian Regularized Neural Network (CO R2:0.71, RMSE: 243.3; H_2S R2:0.75, RMSE: 48.1).

Prediction of CO emission

Figure 1 presents the simulation performed with the chosen models: Generalized Boosted Regression Models, SVM with RBF Kernel, Recursive Partitioning and RPART and Bayesian Regularized Neural Network. These models were compared to empirical data, in that case, it was possible to specify individual models. The characteristic of each model was an increase in CO emissions relative to empirical data. For the empirical data (Fig. 1e), CO emissions were observed to be from 0 to 2126.51 µg CO·g d.m.$^{-1}$ (Supplementary Materials Table S1). The lowest gas emission values of less than 1000 µg CO·g d.m.$^{-1}$ were for materials incubated at 50 °C, in

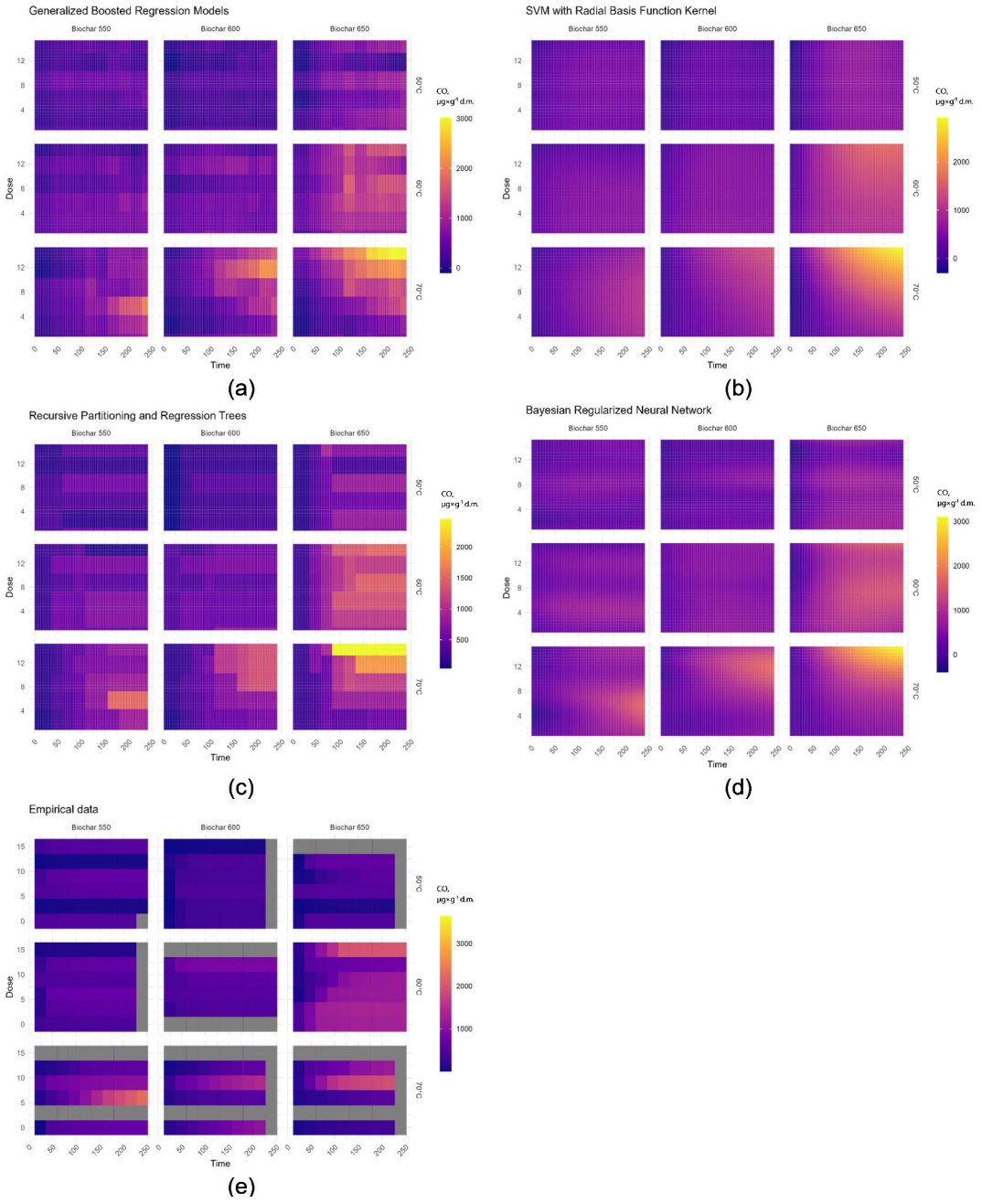

Fig. 1. Predicted CO production (µg CO·g d.m.$^{-1}$) based on biochar temperature production, incubation temperature and dose of biochar, using (**a**) generalized boosted regression models, (**b**) SVM with radial basis function Kernel, (**c**) recursive partitioning and regression trees, (**d**) Bayesian regularized neural network, (**e**) empirical data.

which case the type of biochar did not significantly affect the increase in emissions. Equally low values were seen for material enriched with BC550, while incubated at 60 °C. The highest values, exceeding 2000 µg CO·g d.m.$^{-1}$, were recorded for material with 6% BC550, incubated at 70 °C. High values also characterized the materials with 15% and 9% BC650, incubated at 60 °C and 70 °C, respectively. A characteristic of the models obtained was an overestimation of emissions in areas of missing data present in the empirical data, caused by device failure. The highest emission values were observed for the Bayesian Regularized Neural Network (Fig. 1d) for the material with 15% BC650 incubated at 70 °C (3104.68 µg CO·g d.m.$^{-1}$) (Supplementary Materials Table S1); additionally, this was the model that predicted emission with the highest accuracy. For this model, adding 3 and 6% biochar incubated at 50 °C was the most effective for reducing emissions, irrespective of the pyrolysis process temperature. The Generalized Boosted Regression Model and SVM were the least precise tools for predicting CO emissions (Fig. 1a, b). With this model, particularly for BC650 incubated at 60 °C and 70 °C and for BC600 incubated at 60 °C, a significant overestimation of emissions was observed that was not present in the empirical data. Furthermore, it has been observed that certain models exhibited varying degrees of accuracy in predicting emissions. Additionally, it is essential to note that the selected models may not be suitable for extrapolating data beyond the time range during which measurements were taken.

Prediction of CO_2 emission

Figure 2 shows approximated CO_2 emissions (Fig. 2a–d) and empirical data (Fig. 2e) collected during the laboratory research. In the case of the empirical data, especially for materials with BC600 and BC650 and stored at 70 °C, there was a significant reduction in emissions relative to temperatures of 50 °C and 60 °C; in these cases, emissions reached values close to zero. This indicates that the higher temperatures of the pyrolysis process and the higher storage temperature of the material positively affect the adsorption of CO_2 emissions. BC550 also significantly reduced gas emissions, but only at 6% and 12% doses. A dose of 15% BC650 incubated at 60 °C also effectively reduced CO_2 emissions. The results of the empirical data were in the range 4.58-888.84 mg CO_2·g d.m.$^{-1}$ (Supplementary Materials Table S2), while the highest modeled emission fell for the Bayesian Regularized Neural Network (Fig. 2d) and was embedded in the range 0-620.61 mg CO_2·g d.m.$^{-1}$ (Supplementary Materials Table S2), at the same time, it was the model that performed best in approximating the results from the input data. Equally accurate results were obtained from the RPART model, while in this case there was also a significant underestimation of the predicted final values, as the maximum CO_2 emission was 592.54 mg CO_2·g d.m.$^{-1}$ (Supplementary Materials Table S2) and was observed for material with a 6% dose of BC650 incubated at 60 °C. Despite a relatively high R^2 (> 0.7), the tool with the lowest modeling efficiency for gas emissions was the Generalized Boosted Regression Models model, which underestimated the actual CO_2 production the most of all the models presented graphically.

Prediction of H_2S emission

Figure 3a-d shows a graphical representation of the models predicting the average concentration of H_2S emissions in the test material and contrasts them with the empirical data shown in Fig. 3e. The range of results within which the empirical data fell was from 0.03 µg H_2S·g d.m.$^{-1}$ to 659.44 µg H_2S·g d.m.$^{-1}$ (Supplementary Materials Table S3). The lowest emissions were observed for material incubated at 50 °C; the type of biochar (depending on the temperature of the pyrolysis process) did not have a particularly significant effect on H_2S production. This suggests that H_2S emissions reduction is influenced only by storage conditions, such as lower temperatures, and not by the dose or type of biochar used. The area on the heatmap with the highest gas emissions fell for the 15% BC650 additive stored at 70 °C. Again, the model with the highest performance was the Bayesian Regularized Neural Network, the nature of the prediction in this case was very close to the empirical data, as the range of results obtained was from 0.65 µg H_2S·g d.m.$^{-1}$ to 719.20 µg H_2S·g d.m.$^{-1}$ (Supplementary Materials Table S3), In addition, it was the only model for which $R^2 > 0.7$ was observed. A model with a similar level of fit was RPART, while for the final approximation results a significant under-estimation of emissions was observed relative to the control sample and settled in the range from 11.14 µg H_2S·g d.m.$^{-1}$ to 414.96 µg H_2S·g d.m.$^{-1}$ (Supplementary Materials Table S3). The lowest level of fit of the data to the model was observed for Generalized Boosted Regression Models. A possible reason for the low levels of model fit was the failure to include factors in the input data that directly affect H_2S emissions.

Prediction of NH_3 emission

Figure 4 depicts approximate CO_2 emissions (Fig. 4a–d) and empirical data (Fig. 4e) obtained from laboratory research. The experimental data shown in Fig. 4e had a range from 0.04 µg NH_3·g d.m.$^{-1}$ to 215.58 µg NH_3·g d.m.$^{-1}$ (Supplementary Materials Table S4). The highest emissions were observed for the control sample and the material with 9% BC550 incubated at 60 °C. Material with 15% BC550 addition stored at 50 °C also achieved high emission values. The lowest NH_3 emission values were recorded for BC650, as this type of biochar reduced the measured emissions in the material regardless of dose and storage temperature. Similarly to the other gas emissions, the Bayesian Regularized Neural Network was the most successful model, generating results ranging from 0.04 µg NH_3·g d.m.$^{-1}$ to 252.27 µg NH_3·g d.m.$^{-1}$ (Supplementary Materials Table S4). Despite a slight over-prediction, the model had the highest fit, as evidenced by high R^2 values and low RMSE. RPART (Fig. 4c) was a model with a similar degree of fit; additionally, the approximated values were not as over-predicted as those of the Bayesian Regularized Neural Network (Fig. 4d). The highest emissions predicted by this model were observed for material enriched with doses of biochar 3, 6 and 9% BC550, incubated at 60 °C. A significant reduction in predicted NH_3 emission was present for the SVM with the RBF Kernel model (Fig. 4e). For that model, the maximum approximated emission was less than 200 µg NH_3·g d.m.$^{-1}$ (Supplementary Materials Table S4). In addition, all models failed to cope with data extrapolation beyond the designated time interval.

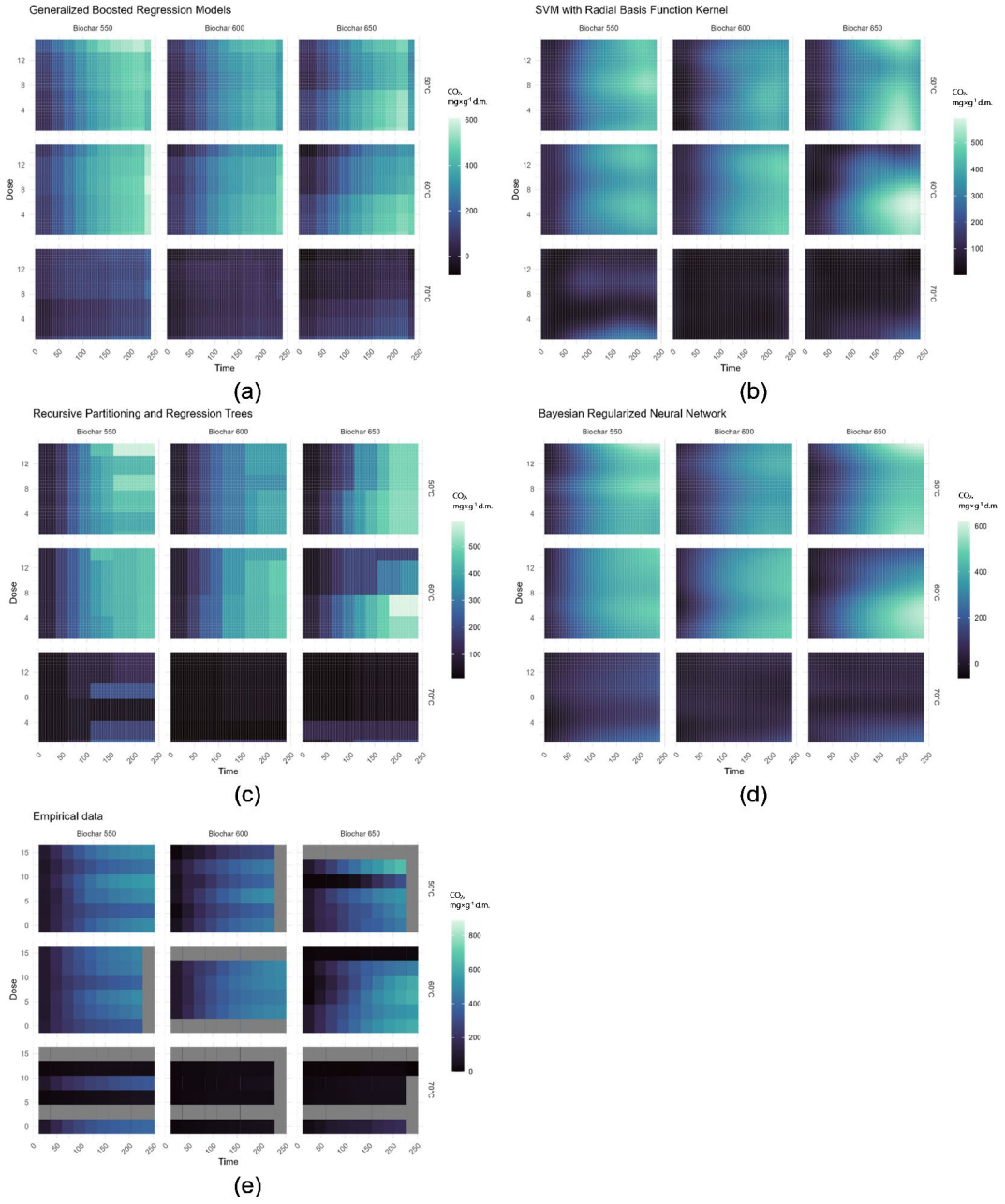

Fig. 2. Predicted CO_2 production (mg CO_2·g d.m.$^{-1}$) based on biochar temperature production, incubation temperature and dose of biochar, using (**a**) generalized boosted regression models, (**b**) SVM with radial basis function Kernel, (**c**) recursive partitioning and regression trees, (**d**) Bayesian regularized neural network, (**e**) empirical data.

The high R^2 and RMSE values for each of the examined models presented graphically demonstrate the high applicability potential of using artificial intelligence to predict NH_3 emissions during the composting process.

Discussion

Improving the efficiency and quality of composting is the primary issue for sustainable composting. Although composting has many advantages in treating organic waste, many problems and challenges are still associated with emissions. Various emissions like NH_3, CO and H_2S, as well as greenhouse gases such as CO_2, and N_2O are generated during the decomposition of organic compounds[18]. It is understood that emissions released during the composting process are influenced by both the feedstock's characteristics and the process's conditions. Effective management emissions techniques such as adsorption/optimizing C/N ratios[19] (for CO_2 reduction), minimizing N losses (for NH_3 reduction)[20], and improving pile oxygenation[21] (for H_2S and CO reduction) can help to control these emissions. One promising approach to enhancing composting conditions to reduce the listed emissions involves using compost's biochar in small quantities[22]. These observations may explain can

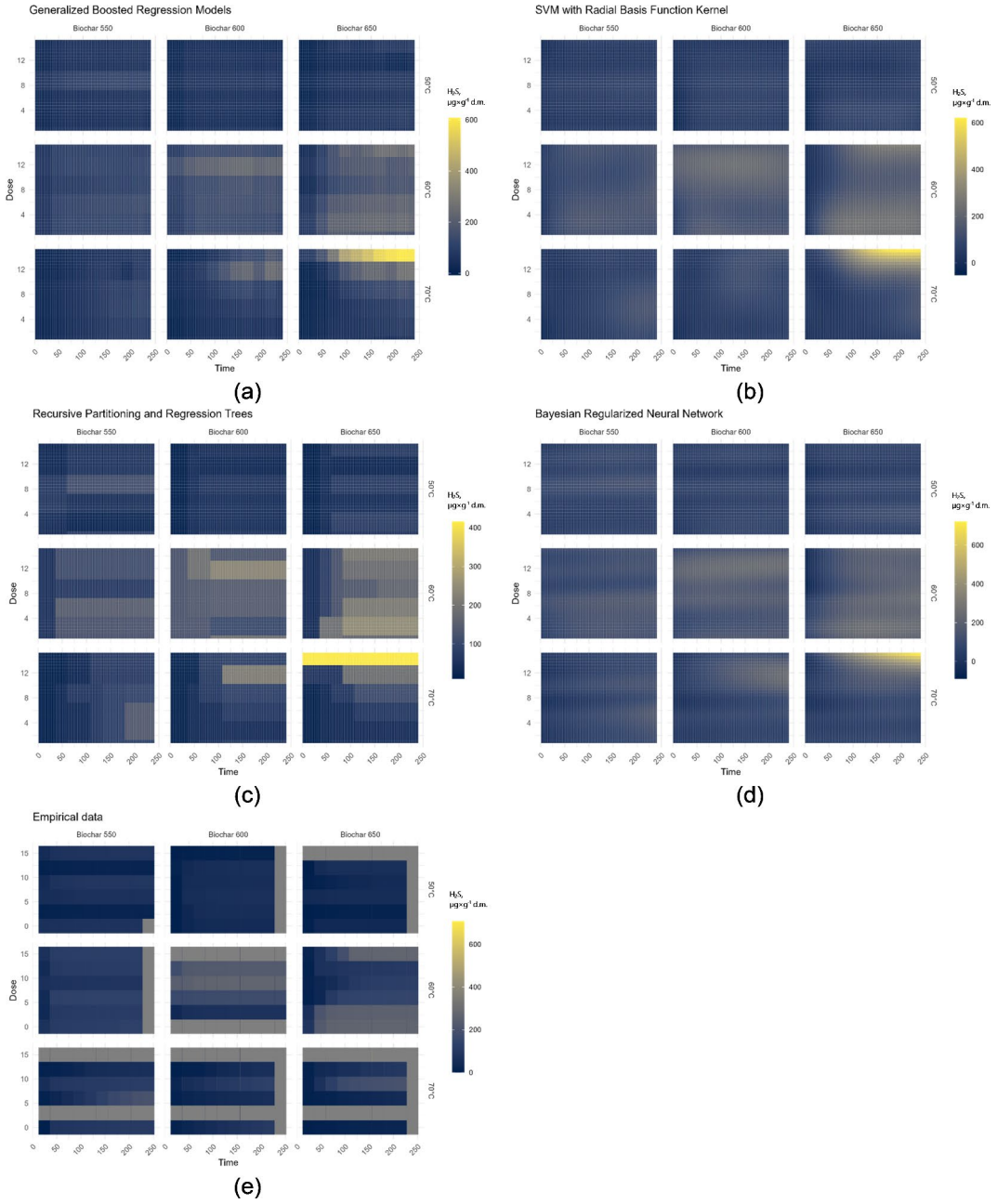

Fig. 3. Predicted H_2S production (μg H_2S·g d.m.$^{-1}$) based on biochar temperature production, incubation temperature and dose of biochar, using (**a**) generalized boosted regression models, (**b**) SVM with radial basis function Kernel, (**c**) recursive partitioning and regression trees, (**d**) Bayesian regularized neural network, (**e**) Empirical data.

observed correlations between the emissions and support the accuracy of emissions modeling based on other emissions used in this study.

The novel analytical methods based on machine learning (ML) models can explore the relationship between different parameters and draw universal conclusions, which was used to predict emissions during green waste composting. Using modeling techniques can significantly decrease costs and expedite implementing new composting practices, especially compared to laboratory and pilot-scale investigations. This makes it an attractive option for exploring innovative composting methods[23].

Currently, ML's research on aerobic composting is still in its early stages. ML models could enhance the initial mixture of biowaste streams and optimal amounts for composting and thereby help to accelerate the process[24], and valuable tool for optimizing process performance in terms of costs, efficiency, and environmental impact by simulating and predicting the process outcome[25]. However mechanism-derived mathematical models may no longer be sufficient, what made the authors focus on predicting the composting process using ML.

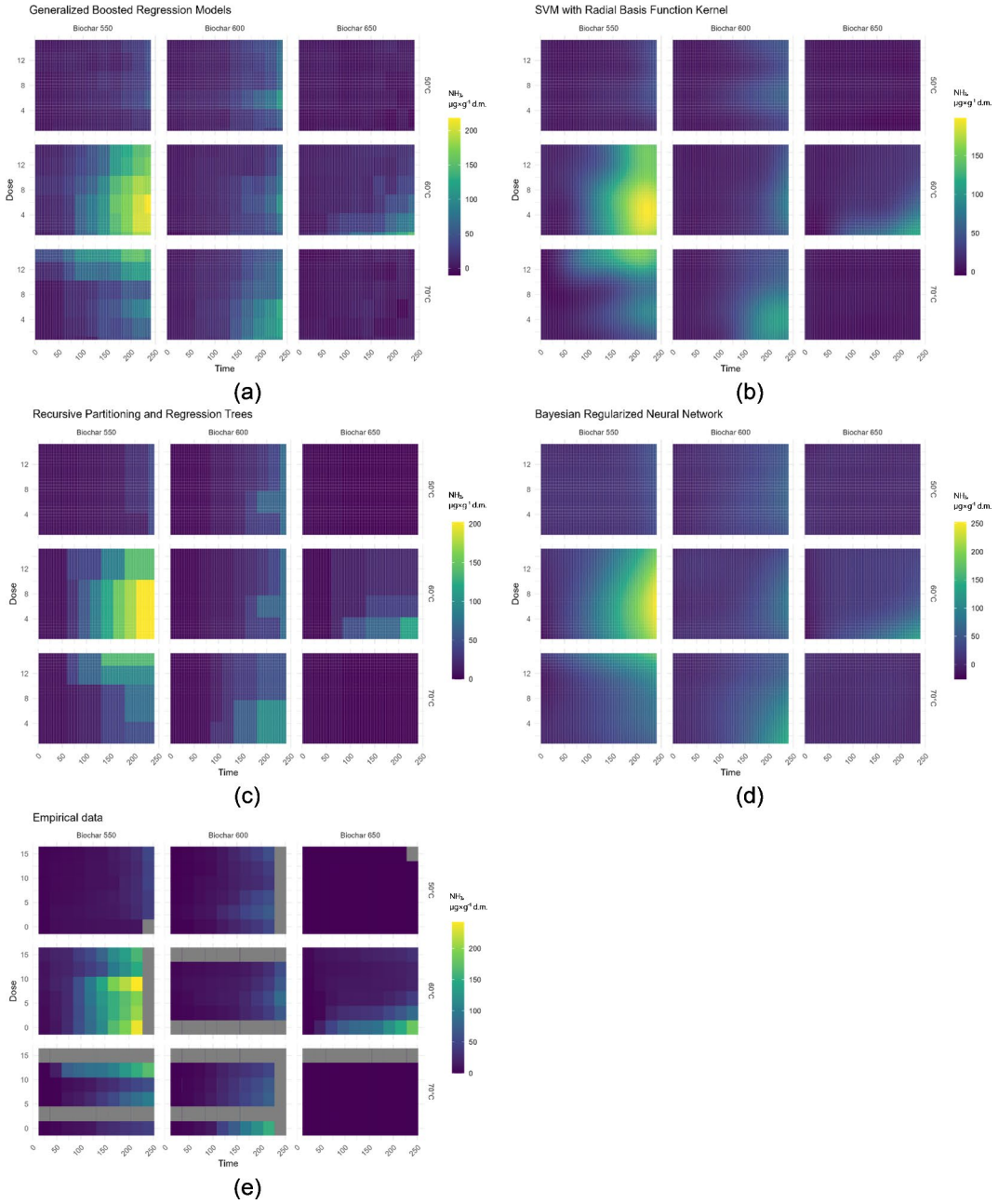

Fig. 4. Predicted NH_3 production (μg NH_3·g d.m. $^{-1}$) based on biochar temperature production, incubation temperature and dose of biochar, using (**a**) generalized boosted regression models, (**b**) SVM with radial basis function Kernel, (**c**) recursive partitioning and regression trees, (**d**) Bayesian regularized neural network, (e) empirical data.

As shown so far, an ML in composting focuses mainly on predicting the compost maturity and compost properties i.e., pH, EC, GI, TN, TOC, etc., with only a few papers concerned with emissions[26]. The accuracy of ML models used in composting process prediction changed in the range of 0.56–0.99 for R^2, but in most cases showed good fit > 0.7. Common ML models used in composting are as follows: Random Forest (RF), Artificial Neural Network (ANN), Support Vector Regression (SVR), Decision Tree, and Decision Support (DS) what it in line with in general biological when the most commonly employed ML algorithms are: ANNs and tree-based models (RF/DT/GBR)[27]. RF and ANN are observed to have the best prediction performance, and the accuracy of R^2 was usually > 0.9. Compared to this study, the best ML models were also ANN (Bayesian Regularized Neural Network), and DT (RPART).

A limited number of authors concentrate on precise forecasts of CO_2 or NH_3 emissions from feedstock composting—furthermore, no research centers on the anticipation of CO or H_2S during composting using machine learning techniques. Li. et al. used various ML models to predict CO_2 emissions based on input variables such as TOC, TN, C/N ratio, cellulose, hemicellulose, and lignin, however those analysis are demanding, time

consuming and costly. The different models had varying levels of RMSE, with AdaBoost at 49.8, Bagging at 80.6, Gradient Boost at 99.9, Random Forest at 83.0, KNN at 55.0, and Decision Tree at 101.8. These results are similar to those presented in this research, as shown in Table 1. Researchers found the highest R^2 score of 0.88 accuracy for Random Forest. Bayesian Regularized Neural Network had the best accuracy of 0.81 in the study, while RF achieved an R^2 score of 0.74 for CO_2 emissions production. This indicates that further research should explore the potential of this type of ML model. In another study for predicting NH_3 emissions during composting sewage sludge with straw, Artificial Neural Network (ANN) was utilized. The ANN achieved an R^2 score of over 0.97 by using temperature, pH, EC, C/N, and N-NH_4 as input parameters[24]. The latest study has also confirmed the efficacy of the XGB model in predicting NH_3 emissions during composting, with an R^2 value of 0.966[4]. The findings of this study suggest that controlling gaseous emissions from green waste composting with compost's biochar can be achieved by monitoring the emissions of other gases e.g. CO_2 output from composting is controllable by CO, H_2S, and NH_3 emissions. In real industrial composting, gas emissions can be accurately forecasted using real-time data from gas sensors with minimal inputs. Previous research indicates that a higher aeration rate can decrease H_2S, NH_3, and CO emissions but can lead to increased energy consumption for the business. Hence, to achieve compost maturity, it is possible to adjust the aeration rate at various stages to minimize NH_3 emissions and enhance aeration efficiency. Additionally, other studies have also shown that the modeling observations are similar even when other input materials were used for the composting process (organic waste and kitchen waste)[14]. This requires confirmation and adjustment of the models, but gives a perspective for use in different types of composting.

It is important to note that the experimental data used in this study are based on the observations from previous publications and may not fully reflect the control of CO, CO_2, H_2S and NH_3 emissions from composting. Nevertheless, this solution can provide valuable insights for future studies and practices with a larger dataset (especially collected in field study) and more sophisticated ML techniques.

The article underscores the potential of machine learning (ML) in predicting gas emissions. However, given the high level of innovation and complexity in the composting process, further research and development in this technology, particularly at an industrial scale, is essential. Firstly, it is crucial to compile a larger dataset of composting data to improve the reliability of prediction results. Secondly, additional ML models should be explore to forecast emissions based on specific composting feedstock characteristics. Finally, it is advisable to utilize practical composting data to train ML models for real-world applications and to enhance composting efficiency.

Conclusions

This study utilized machine learning (ML) models to predict the emissions (CO, CO_2, H_2S, NH_3) during the first 10 days of composting with compost's biochar addition. The ML models to predict CO and H_2S during composting were demonstrated. ML models such as acritical neural network (ANN) and decision tree (DT) have demonstrated satisfactory results. A quality assessment of the developed ML models has shown that the best predictive capacity was reached for ANN (Bayesian Regularized Neural Network; R^2 accuracy CO:0,71, CO_2:0,81, NH_3:0,95, H2S:0,72) and DT (RPART; R^2 accuracy CO:0,69, CO_2:0,80, NH_3:0,93, H_2S:0,65). The practical implications of the predictive model reveal that the estimated CO, CO_2, H_2S and NH_3 values closely align with the actual values observed during real composting. The study confirms that the use of AI for optimizations of emissions during composting has good potential and can be used to improve the safety and effectiveness of the process.

Materials and methods

The experiment design and procedure

The machine learning model training (Sect. "Data pre-processing" and "Selection ML model selection and training machine learning algorithms evaluation") relied on data from published sources[22]. The study centers on the influence of compost's biochar (BC) addition to feedstock and how it impacts CO, CO_2, H_2S, and NH_3 emissions during the early stages of laboratory composting. The presence of these gases presents a potential hazard to the personnel employed at the composting facility, as well as a risk of environmental harm. The composting experiments used a feedstock mix of 90% green waste and 10% sewage sludge acquired from a composting plant (Best-Eko, Rybnik, Poland). The compost's biochars (BC550; BC600; BC650), produced at different pyrolysis temperatures, were applied at doses of 0, 3, 6, 9, 12 and 15% d.m as shown in Fig. 5. The biochars were produced from fully mature certified compost - BEST-TERRA.The specific surface area of tested biochars reached: BC550: 6,1 $m^2 \cdot g^{-1}$, BC600: 29,3 $m^2 \cdot g^{-1}$ and BC650: 39,2 $m^2 \cdot g^{-1}$. The average biochars' pore size decreased as the temperature of the pyrolysis process increased from 2,2 nm (BC550) to 1,4 nm (BC650). The appropriate biochar variant was added to the feedstock, placed in 1 L reactors, and kept at 50, 60, or 70 °C in a thermostatic cabinet for 10 days to simulate the early-stage composting process conditions. Due to the challenges of maintaining optimal temperature conditions for composting in a laboratory setting, the initial intensive phase of composting typically involves selecting from three commonly observed temperature ranges: 50, 60, and 70 °C. Our previous study[22] shows, the initial 10 days are critical in determining a substantial portion of the total emissions and can significantly impact emissions in the later stages of composting. Therefore, this study meticulously examines the emissions during this initial stage. The concentrations of CO, CO_2, NH_3 and H_2S were measured daily throughout the composting process and then used to calculate emissions. The comprehensive composting process protocol for this study was outlined in previous research[22].

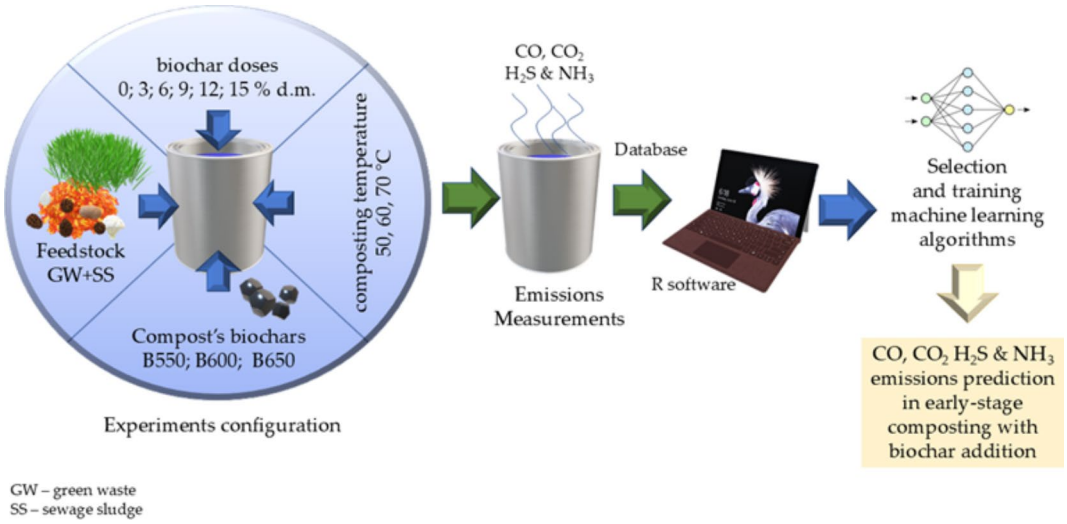

Fig. 5. Experiments configurations.

Gas production monitoring

During the laboratory composting, everyday gas concentrations of CO, CO_2, NH_3 and H_2S were done. A portable electrochemical gas analyzer was used for gas concentration measurements (Nanosens DP-28 BIO; Wysogotowo, Poland). Concentrations of CO, H_2S, and NH_3 were determined in ppm in the following ranges: CO 0–2000 ppm (± 20 ppm), H_2S, NH_3 0–1000 ppm (± 10 ppm). CO_2 was specified in percentages in the range of 0–100% (± 2%). Each measurement lasted 45 s, followed by automatic cleaning of the analyzer. Equations 1 to 4 show the calculation scheme used when calculating the conversion of gas concentrations to emissions:

- Concentration-to-volume conversion for H_2S, CO and NH_3 (ppm):

$$V = \frac{1000 \cdot M \cdot \left(1{,}66 \cdot 10^{-24}\right) \cdot \left(\left(2{,}68839 \cdot 10^{22}\right) \cdot a\right)}{1000000} \tag{1}$$

where: V – gas volume, m^3, M – molar mass, mol, a – gas concentration, ppm.

- Concentration-to-volume conversion for CO_2 (%):

$$V = \frac{1000 \cdot M \cdot \left(1{,}66 \cdot 10^{-24}\right) \cdot \left(\left(2{,}68839 \cdot 10^{22}\right) \cdot a\right)}{100} \tag{2}$$

- Emission (H_2S, CO and NH_3):

$$E = \frac{V}{d.m. \cdot 1000} \tag{3}$$

where: E – emission, μg·g $s.m.^{-1}$, d.m. – dry mass, g.

- Emission (CO_2):

$$E = \frac{V}{d.m.} \tag{4}$$

Data pre-processing

Figure 6 depicts the data processing steps. Initially, 66,048 datasets (hourly measurement of CO, CO_2, NH_3 and H_2S emission) were extracted from the selected references without missing data. Subsequently, the collected data was normalized from 0 to 1 using Z-Score normalization. Finally, the dataset was randomly divided into training and testing datasets to enhance prediction accuracy, as previously reported[11]. The data was split into training/validation/test groups in a 70%/15%/15% proportion. For the fine-tuning process, k-fold cross-validation with grid search was employed. The training dataset assisted in adjusting the hyperparameters and enhancing the prediction abilities of the model, while the testing dataset was used to evaluate the model's performance and select the appropriate model by comparing the RMSE and R^2 values[28].

$$R^2 = 1 - \left[\frac{\sum_{t=1}^{T}(y*_t - y_t)^2}{\sum_{t=1}^{T} y*_t - y_t{}^2}\right] \tag{5}$$

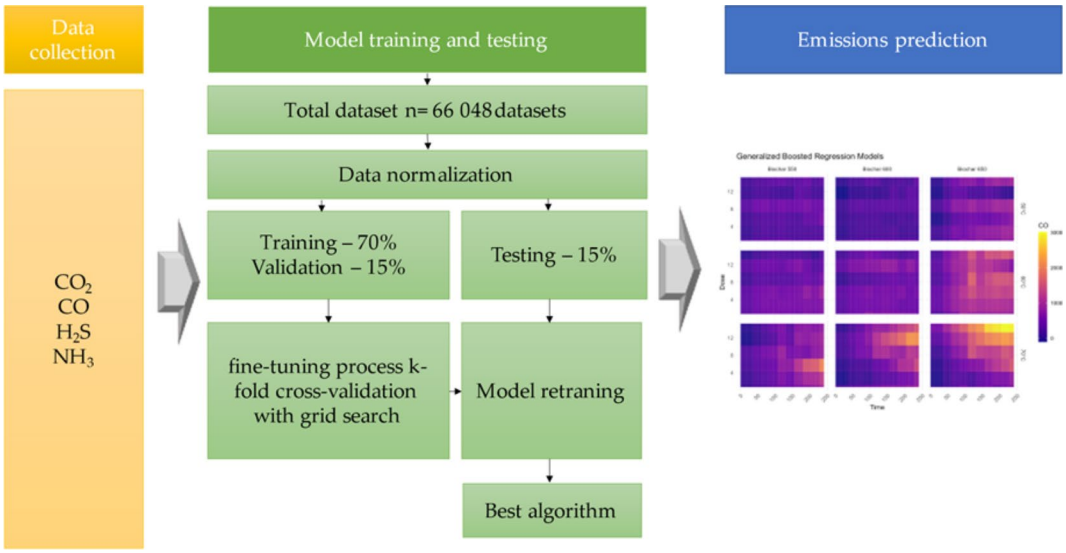

Fig. 6. Machine learning flowchart for predicting emissions from composting with biochar addition.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(y*_t - y_t)^2}{T}} \tag{6}$$

Selection ML model selection and training machine learning algorithms evaluation

In this study, ten learning algorithms were evaluated, including both machine set learning and non-set learning. To assess the viability of machine learning methods in the prediction of CO, CO_2, NH_3 and H_2S emissions during the first stage of composting, various classes of methods were compared: Linear Models, Tree-Based Models (also part of Ensemble Methods), Support Vector Machines (SVM) and Neural Networks. Calculations were performed using R for Windows[29] (ver 4.3.2, Vienna, Austria) with caret[30] and h2o[31] libraries. The data used for model training related to CO, CO_2, NH_3 and H_2S emissions from composting were obtained from published studies. To predict each gas emission (CO, CO_2, NH_3 and H_2S) individually, principal component analysis (PCA) was conducted to exclude irrelevant parameters. The PCA analysis indicated that observed emissions have a significant correlation. The use of other parameters is not justified. PCA (a linear dimensionality reduction algorithm) facilitated dimensions standardization and reduction of the initial complexity of the model. Moreover, it will be easier to apply the model in practice if the variables are limited to those that can be easily and cheaply implemented in composting i.e. gas emissions (Supplementary Materials Figure S1). In model training and prediction, the output and input of the model were the data about CO, CO_2, NH_3 and H_2S emissions. During the training, the data about the other emissions were utilized as input when one gas emission was used as an output.

The top four models (Generalized Boosted Regression Models (GBM); SVM with Radial Basis Function (RBF) Kernel Nearest Neighbor Models; Bayesian Regularized Neural Network; Recursive Partitioning and Regression Trees) were depicted as heatmaps, revealing the impact of the four variables: biochar dose, biochar type, incubation temperature, and time on gas emission. Finally, the predicted emissions were compared to the actual emissions to determine the models' accuracy.

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Received: 18 September 2024; Accepted: 5 November 2024
Published online: 08 November 2024

References

1. Bao, M. et al. Greenhouse gas emission during swine manure aerobic composting: Insight from the dissolved organic matter associated microbial community succession. *Bioresour Technol.* **373**, 128729 (2023).
2. Jiang, L. et al. Adding siderophores: A new strategy to reduce greenhouse gas emissions in composting. *Bioresour Technol.* **384**, 129319 (2023).
3. Zhou, Y. et al. Synergistic effects of biochar derived from different sources on greenhouse gas emissions and microplastics mitigation during sewage sludge composting. *Bioresour Technol.* **387**, 129556 (2023).
4. Wang, B., Zhang, P., Qi, X., Li, G. & Zhang, J. Predicting ammonia emissions and global warming potential in composting by machine learning. *Bioresour Technol.* **411**, 131335 (2024).
5. Sobieraj, K., Stegenta-Dąbrowska, S., Zafiu, C., Binner, E. & Białowiec, A. Carbon monoxide production during bio-waste composting under different temperature and aeration regimes. *Materials*. **16**, 4551 (2023).

6. Tran, H. T. et al. Succession of biochar addition for soil amendment and contaminants remediation during co-composting: A state of art review. *J. Environ. Manage.* **342**, 118191 (2023).
7. Sadegh, F., Sadegh, N., Wongniramaikul, W., Apiratikul, R. & Choodum, A. Adsorption of volatile organic compounds on biochar: A review. *Process Saf. Environ. Prot.* **182**, 559–578 (2024).
8. Dang, B. T. et al. Current application of seaweed waste for composting and biochar: A review. *Bioresour Technol.* **375**, 128830 (2023).
9. Ye, Z. et al. Tackling environmental challenges in pollution controls using artificial intelligence: A review. *Sci. Total Environ.* **699**, 134279 (2020).
10. Zhong, S. et al. Machine learning: New ideas and tools in environmental science and engineering. *Environ. Sci. Technol.* **55**, 12741–12754 (2021).
11. Wang, N. et al. Predicting maturity and identifying key factors in organic waste composting using machine learning models. *Bioresour Technol.* **400**, 130663 (2024).
12. Singh, T. & Uppaluri, R. V. S. Application of ANN and traditional ML algorithms in modelling compost production under different climatic conditions. *Neural Comput. Appl.* **35**, 13465–13484 (2023).
13. Boniecki, P., Dach, J., Pilarski, K. & Piekarska-Boniecka, H. Artificial neural networks for modeling ammonia emissions released from sewage sludge composting. *Atmos. Environ.* **57**, 49–54 (2012).
14. Ding, S. et al. Improving kitchen waste composting maturity by optimizing the processing parameters based on machine learning model. *Bioresour Technol.* **360**, 127606 (2022).
15. Li, Y., Li, S., Sun, X. & Hao, D. Prediction of carbon dioxide production from green waste composting and identification of critical factors using machine learning algorithms. *Bioresour Technol.* **360**, 127587 (2022).
16. Xie, Q., Ni, J. & Su, Z. A prediction model of ammonia emission from a fattening pig room based on the indoor concentration using adaptive neuro fuzzy inference system. *J. Hazard. Mater.* **325**, 301–309 (2017).
17. Küçüktopcu, E. & Cemek, B. Comparison of neuro-fuzzy and neural networks techniques for estimating ammonia concentration in poultry farms. *J. Environ. Chem. Eng.* **9**, 105699 (2021).
18. Andraskar, J., Yadav, S. & Kapley, A. Challenges and control strategies of odor emission from composting operation. *Appl. Biochem. Biotechnol. 2021.* **193:7 193**, 2331–2356 (2021).
19. Li, H., Zhang, T., Tsang, D. C. W. & Li, G. Effects of external additives: Biochar, bentonite, phosphate, on co-composting for swine manure and corn straw. *Chemosphere*. **248**, 125927 (2020).
20. Awasthi, M. K., Duan, Y., Awasthi, S. K., Liu, T. & Zhang, Z. Influence of bamboo biochar on mitigating greenhouse gas emissions and nitrogen loss during poultry manure composting. *Bioresour Technol.* **303**, 122952 (2020).
21. Sobieraj, K., Stegenta-Dąbrowska, S., Koziel, J. A. & Białowiec, A. Modeling of CO accumulation in the headspace of the bioreactor during organic waste composting. *Energies* . **14**, 1367 (2021).
22. Stegenta-Dąbrowska, S., Syguła, E., Bednik, M. & Rosik, J. Effective carbon dioxide mitigation and improvement of compost nutrients with the Use of composts' Biochar. *Materials*. **17**, 563 (2024).
23. Kabak, E. T., Yolcu, C., Aydın Temel, O., Turan, N. G. & F. & Prediction and optimization of nitrogen losses in co-composting process by using a hybrid cascaded prediction model and genetic algorithm. *Chem. Eng. J.* **437**, 135499 (2022).
24. Li, Y., Xue, Z., Li, S., Sun, X. & Hao, D. Prediction of composting maturity and identification of critical parameters for green waste compost using machine learning. *Bioresour Technol.* **385**, 129444 (2023).
25. Walling, E., Trémier, A. & Vaneeckhaute, C. A review of mathematical models for composting. *Waste Manage.* **113**, 379–394 (2020).
26. Aydın Temel, F., Cagcag Yolcu, O. & Turan, N. G. Artificial intelligence and machine learning approaches in composting process: A review. *Bioresour Technol.* **370**, 128539 (2023).
27. Chen, L. et al. Applications of machine learning tools for biological treatment of organic wastes: Perspectives and challenges. *Circular Econ.* **3**, 100088 (2024).
28. Abdi, R., Shahgholi, G., Sharabiani, V. R., Fanaei, A. R. & Szymanek, M. Prediction compost criteria of organic wastes with biochar additive in in-vessel composting machine using ANFIS and ANN methods. *Energy Rep.* **9**, 1684–1695 (2023).
29. R Core Team. R: A Language and environment for statistical computing. Preprint at. (2023).
30. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).
31. Fryda, T., LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., Kraljevic, T., Nykodym, T., Aboyoun, P., Kurka, M., Malohlava, M., Poirier, S. & Wong, W. R Interface for the 'H2O' Scalable Machine Learning Platform_. R package version 3.44.0.3. Preprint at (2024).

Acknowledgements

The authors would like to thank the Best-Eko sp. z.o.o. company for the opportunity to collect compost and for cooperation.

Author contributions

Conceptualization, S.S-D. and J.R.; methodology, S.S-D., J.R. and M.K.; software, M.K.; validation, S.S-D., J.R. and M.K.; formal analysis, S.S-D., J.R. and M.K.; investigation, S.S-D., J.R. and M.K.; resources, S.S-D. and J.R.; data curation, S.S-D. and J.R.; writing—original draft preparation, S.S-D. and J.R.; writing—review and editing, S.S-D. and J.R.; visualization, S.S-D., J.R. and M.K.; supervision, S.S-D. and J.R.; project administration, S.S-D.; funding acquisition, S.S-D. All authors have read and agreed to the published version of the manuscript.

Funding

The research leading to these results has received funding from the Norway Grants 2014–2021 through the National Centre for Research and Development. Project number: NOR/SGS/CompoChar/0090/2020. The APC is financed by Wrocław University of Environmental and Life Sciences.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-79010-0.

Correspondence and requests for materials should be addressed to S.S.-D.