Earth Systems and Environment (2024) 8:1109–1125
https://doi.org/10.1007/s41748-024-00424-x

ORIGINAL ARTICLE

Assessment of Advanced Machine and Deep Learning Approaches for Predicting CO_2 Emissions from Agricultural Lands: Insights Across Diverse Agroclimatic Zones

Endre Harsányi[1,2] · **Morad Mirzaei**[3,4] · **Sana Arshad**[5] · **Firas Alsilibe**[6] · **Atilla Vad**[2] · **Adrian Nagy**[9] · **Tamás Ratonyi**[1] · **Manouchehr Gorji**[7] · **Main Al-Dalahme**[8] · **Safwan Mohammed**[1,2]

Received: 22 February 2024 / Revised: 21 May 2024 / Accepted: 21 June 2024 / Published online: 3 July 2024
© The Author(s) 2024

Abstract

Prediction of carbon dioxide (CO_2) emissions from agricultural soil is vital for efficient and strategic mitigating practices and achieving climate smart agriculture. This study aimed to evaluate the ability of two machine learning algorithms [gradient boosting regression (GBR), support vector regression (SVR)], and two deep learning algorithms [feedforward neural network (FNN) and convolutional neural network (CNN)] in predicting CO_2 emissions from Maize fields in two agroclimatic regions i.e., continental (Debrecen-Hungary), and semi-arid (Karaj-Iran). This research developed three scenarios for predicting CO_2. Each scenario is developed by a combination between input variables [i.e., soil temperature (Δ), soil moisture (θ), date of measurement (SD), soil management (SM)] [i.e., SC1: (SM + Δ + θ), SC2: (SM + Δ), SC3: (SM + θ)]. Results showed that the average CO_2 emission from Debrecen was 138.78 ± 72.04 ppm (n = 36), while the average from Karaj was 478.98 ± 174.22 ppm (n = 36). Performance evaluation results of train set revealed that high prediction accuracy is achieved by GBR in SC1 with the highest R^2 = 0.8778, and lowest root mean squared error (RMSE) = 72.05, followed by GBR in SC3. Overall, the performance MDLM is ranked as GBR > FNN > CNN > SVR. In testing phase, the highest prediction accuracy was achieved by FNN in SC1 with R^2 = 0.918, and RMSE = 67.75, followed by FNN in SC3, and GBR in SC1 (R^2 = 0.887, RMSE = 79.881). The performance of MDLM ranked as FNN > GRB > CNN > SVR. The findings of the research provide insights into agricultural management strategies, enabling stakeholders to work towards a more sustainable and climate-resilient future in agriculture.

Keywords Carbon cycle · Climate-resilient · Soil management · Agroclimatic regions · Climate smart agriculture

✉ Safwan Mohammed
safwan@agr.unideb.hu

[1] Institute of Land Use, Engineering and Precision Farming Technology, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, Böszörményi 138, Debrecen 4032, Hungary

[2] Institutes for Agricultural Research and Educational Farm, University of Debrecen, Böszörményi 138, Debrecen 4032, Hungary

[3] School of Natural Sciences, Botany Discipline, Trinity College Dublin, Dublin 2, Ireland

[4] Teagasc Crops, Environment and Land Use Programme, Johnstown Castle, Co. Wexford Y35 TC97, Ireland

[5] Department of Geography, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan

[6] Department of Transport Infrastructure and Water Resources Engineering, Széchenyi István University, Egyetem Tér 1, Győr 9026, Hungary

[7] Department of Soil Science and Engineering, Faculty of Agricultural Engineering and Technology, University of Tehran, Karaj, Iran

[8] Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman, Jordan

[9] Faculty of Economics and Business, Institute of Applied Economics, University of Debrecen, Debrecen, Hungary

1 Introduction

Emissions of carbon dioxide (CO_2) as a main greenhouse gases (GHGs) have been significantly increased up to 40% since pre-industrial time attributed to anthropogenic activities and fossil fuel combustion (Mbow et al. 2017). Extensive farming activities like overgrazing, intensive tillage, and improper use of chemical fertilizers and manure lead to emissions of GHGs (Ray et al. 2020; Zaman et al. 2021). Forest or grasslands conversions to arable land remained one the biggest cause of soil carbon release in atmosphere since late seventeenth century. Organic matter decomposition and reduction in plant biomass reduced the soil carbon content and released 1.2—4.6 Gt C/year in past decades (Mitchell et al. 2019; Sauerbeck 2001). Enteric fermentation, pasture manure, biomass burning, and synthetic fertilizers has major contribution of CO_2 emissions from agriculture perspective (Campbell et al. 2014; Sainju et al. 2008; Tubiello et al. 2013). Various climatic factors, such as temperature, rainfall and soil moisture also alter the soil microbial activities and increasing root respiration and resulting in elevated soil CO_2 emissions (Bååth 2018; Mirzai et al. 2022; Ray et al. 2020; Zechmeister-Boltenstern et al. 2018). Hence, soil respiration is the largest contributor to carbon (C) emissions from terrestrial ecosystems (Almagro et al. 2013). Depending on the soil management (SM) and the practices implemented, soil can acts both as either a sink or source of CO_2 emissions (Shakoor et al. 2021).

Different soil cultivation methods, including tillage or non-tillage, have varying impacts on CO_2 emissions. For instance, reduced or no tillage practices contribute to lowering soil temperature and restricting CO_2 emissions (Li et al. 2022; Mancinelli et al. 2023).

On contrary, conventional tillage accelerates the soil microbial activity and decomposition of organic matter that is more relatable with increased CO_2 emissions along with other GHGs (Bhattacharyya et al. 2022; Devkota et al. 2022). While conservation tillage is noted for carbon sequestration and preventing CO_2 emissions by leaving sufficient crop residues covering the subsurface soil (Devkota et al. 2022; Fageria and Moreira 2011; Mitchell et al. 2019). Hence, dynamics of GHGs emissions from agriculture soil are vital to monitor for efficient and strategic mitigating practices and achieving climate smart agriculture (Devkota et al. 2022).

Experimental methods for quantifying of CO_2 emissions faces several constraints including the long-term monitoring, labor requirements, and extensive analysis (Yadav and Wang 2017).Simulating GHGs emissions incorporating various factors like fertilization, climate, soil cultivation practices, and many other factors have been done in numerous research (Li et al. 2010b, 2016; Lloyd et al. 2019; Yu et al. 2013). Previously, several biophysical models such as Root zone water quality model (RZWQM) (Jiang et al. 2018), DayCent (Necpálová et al. 2015), CERES-EGO (Durandeau et al. 2010) DNDC (Li et al. 2010a; Yadav and Wang 2017) had been employed to model the GHGs emissions in different scenarios bounded with some limitations. Main constraints in utilizing biophysical or process based models included large data input, high agro-environmental expertise, complex procedures of model calibration and validation (Smith et al. 2010).

Several Machine and deep learning (ML and DL) methods provide a smart alternative to biophysical methods for various environmental phenomenon such as drought and flood prediction (Hu et al. 2019; Mohammed et al. 2022a), predicting ground water level and agricultural water quality (Mohammed et al. 2024; Singh et al. 2024). Classical ML models such as random forest (RF), support vector machine (SVM), gradient boosting regression (GBR), and least absolute shrinkage and selection operator (LASSO) have been used for prediction modelling of GHGs in many regions (Table 1) (Abbasi et al. 2021; Hamrani et al. 2020; Saha et al. 2021; Yan et al. 2020; Zhou et al. 2022).

Deep learning also involves utilization of more effective algorithms such as artificial neural networks (ANN) and convolution neural networks used for (CNN) for complex modelling (LeCun et al. 2015; Reichstein et al. 2019) and predicting non-linear complex relationships (Altikat 2021; Madu et al. 2017). Exceptional capabilities of NN like one input–output communication, quick response to data noise, more flexibility and adaptability, error tolerance and robust non-linear predictions make them advantageous of classical ML algorithms (Shabani et al. 2021). Wide applications of NN in various environmental studies such as soil erosion (Gholami et al. 2018), nutrients runoff (Kim and Gilley 2008), heavy metals in soil (Naderi et al. 2017), groundwater levels (Iqbal et al. 2020), and hydrological modelling (Pradhan et al. 2020) make it a good choice to be used for modelling GHGs emissions from various sectors (Ganzenmüller et al. 2019). It is also being employed for prediction modelling of CO_2 emission from soil and land use land cover in many countries such as Iran (Altikat 2021; Shabani et al. 2021), Brazil (Freitas et al. 2018; Vitória et al. 2022), Turkey (Yılmaz and Bilgili 2018) and Bangladesh (Fattah et al. 2021). Noticeably, the analysis of the available research published in the Web of Science database revealed that machine and deep learning (green line) is associated with published research related to CO_2 emissions (Fig. 1). However, previous studies on CO_2 emissions did not account for multiple factors in different input combinations or scenarios in varying climatic conditions. In addition, applicability of machine and deep learning algorithms for predicting CO_2 emissions also needs to be tested in different

Table 1 Review of prediction of some GHGs emissions studies using machine and deep learning algorithms in recent years

Prediction Objective	Predictors	ML algorithm	Outperformance	References
Soil GHGs CO_2, CH_4, and N_2O	Micro and neoplastic material	Automated machine learning	Autogluon framework provided high accurate predictions	Lin et al. (2024)
Soil CO_2 emissions in forests ecosystem	Soil and meteorological	GRNN, RBFNN, MLPNN, ANFIS, RF	RF outperformed other models with $R^2=0.70$	Canteral et al. (2023)
Soil GHG emissions CO_2 fluxes	Soil moisture & temp, Soil type, crop type, fertilization, air temp	KNN, SVR, RF, GBR	GBR with $R^2=0.88$	Adjuik and Davis (2022)
Agriculture GHG emissions CO_2 and N_2O	Meteorological, soil temp, soil moisture	Linear regression, cubist, Random Forest, SVM	SVM with non-linear kernels outperformed with and without cover cropping	Singh and Kumar (2022)
CO_2 flux from soil	Crop types, soil temperature and moisture, photosynthetic active radiation, oxygen exchange of soil	ANN and DLNN	DLNN outperformed with 98% accuracy followed by sensitivity analysis	Altikat (2021)
CO_2 and CH_4 emissions	Soil temperature, moisture, humidity, pH, type, and meteorological	Decision tree, DNN	Deep neural network outperformed	Kosamkar and Kulkarni (2021b)
CO_2, NO, and NO_x emissions	Crop residue rates, soil temperature, and soil water filled pore space (WFPS)	RFR, MARS, and GLM	RFR	Mirzaei et al. (2023)

scenarios and environmental settings for a wider applicability. Hence, keeping in view the significant application of advanced machine learning for GHGs predictions, our study aimed to evaluate the ability of two machine learning algorithms namely GBR, SVR, and two deep learning algorithms namely FNN and CNN. This evaluation is conducted across three distinct scenarios of input combinations to predict CO_2 emissions from Maize fields in two agroclimatic regions, i.e., continental (Debrecen-Hungary) and semi-arid (Karaj-Iran).

2 Method

2.1 Study Area and Data Collection (Observation Stage)

CO_2 gas samples were collected from maize (*Zea mays* L.) fields in two different agroclimatic regions. The first one was at the Agriculture Research Station of the College of Agriculture and Natural Resources, University of Tehran, Karaj (Iran) (50°58′ E, 35°48′ N), representing a semi-arid climate. The second was a continental climate located at the Látókép research station at the University of Debrecen (Hungary) (47°33′ E, 21°26′ N). The experimental design included two different soil management practices: (1) conventional tillage and (2) non-tillage. In Karaj, gas samples were collected using the static closed chamber methodology, one of the most used and cost-effective techniques for measuring GHG emissions from agroecosystems. The CO_2 concentration was measured using gas chromatography (Teif Gostar Faraz, TG 2552), equipped with a thermal conductivity detector (TCD).

In Debrecen, CO_2 emission was measured using the digital meter Testo 535 (TESTO; 0560 5350) which measures CO_2 concentration via infrared absorption with 1 ppm of CO_2 (Törő et al. 2019). The average CO_2 emission from Debrecen was 138.78 ± 72.04 ppm ($n = 36$), while the average from Karaj was 478.98 ± 174.22 ppm ($n = 36$) (Table 2). More detailed information about the experimental design and the interaction between different environmental variables (i.e., soil temperature (Δ), and soil moisture (θ)) and CO_2 emission could be retrieved from Mohammed et al. (2022b). Observed CO_2 emissions from both experimental sites were collectively taken as a response or input variable to be predicted in ANN modelling architectures.

2.2 Machine Learning Modelling

The study employed two classical machine learning models, GBR and SVR, alongside two deep learning models, namely

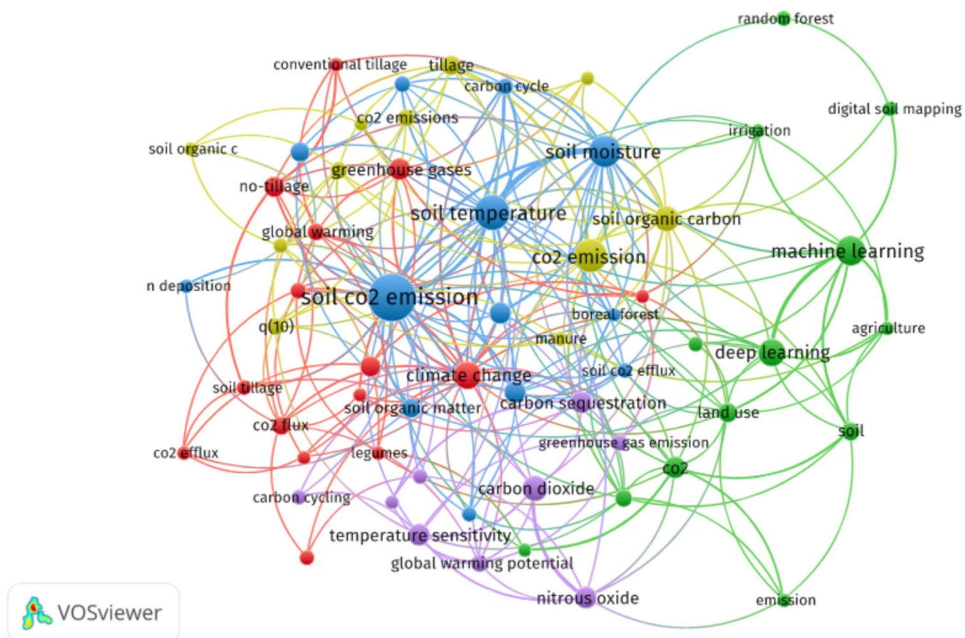

Fig. 1 Bibliometric analysis of research published in the Web of Science database related to machine learning and CO_2 emissions, using the search equation: ("CO_2 emissions") AND ("artificial intelligence" OR "deep learning") AND (soil) AND (climate OR "climate change"). This analysis revealed 550 published research articles and identified 5 clusters

FNN and CNN. This analysis was conducted under three different scenarios, incorporating meteorological and soil management factors (Table 3).

The general principle of machine learning modelling is based on a test size split of 0.25 (75% training and 25% testing). The hyperparameters of machine learning models are tuned using Randomized Search CV function from scikit-learn with 10 iterations and five-fold cross validation to identify the optimal model. The optimal selection of deep learning models is determined through manual adjustments of learning rates, neurons, and activation function. All models are run in python 3, leveraging the computational capabilities of T4GPU in 32 GB RAM for enhanced training and prediction. The hyperparameters details of all models are presented in Table 4.

2.2.1 Gradient Boosting Regression

The gradient boosting regression (GBR) is based on the boosting algorithm encompassed by gradient descent, where subsequent trees are fitted to the residuals of the preceding model, gradually reducing the overall prediction error, and boosting functions widely applied in regressions and classifications problems (Jerome 2001). The iterative process of GBR enhances the predictions and improves the performance of the model. Unlike random forest, which trains multiple decision trees independently and averages their predictions, GBR constructs a series of models sequentially, with each subsequent model focusing on the errors made by its predecessors (Arshad et al. 2023b). This sequential learning process enables Gradient Boosting to correct earlier mistakes effectively, leading to superior predictive accuracy (Otchere et al. 2022). Additionally, it inherently deals with heterogeneous data types and automatically selects relevant features during the process of model training. This ability of selecting appropriate features not only simplifies the modeling process but also enhances the interpretability of model achieving accurate predictions (Wen et al. 2021). Currently, GBR is initialized on standardized features (meteorological and soil management) using randomized search hyperparameter tuning with five-fold cross validation (Table 4). The best fit model is trained and implemented to achieve high prediction accuracy.

2.2.2 Support Vector Regression

Support vector regression (SVR) is a supervised machine learning method that is based on statistical learning approach derived from support vector machines (SVM), which aims to identify the hyperplane that best fits the train data with maximizing the margin between hyperplane and nearest data points, called support vectors (Zhang and O'Donnell 2020).

SVR develops an optimization problem to acquire a regression function that connects input predictor variables to observed response values. Nevertheless, SVR offers a robust and adaptable approach to regression tasks, particularly in situations characterized by non-linear relationships and a tolerance for forecast errors (Arshad et al. 2023a; Awad and Khanna 2015). Currently, hyperparameter tuning of SVR is performed using Randomized Search to explore the best regularization parameter (c), kernel type (linear, and rbf), and kernel coefficient for the radial basis function (gamma) (Table 4).

Table 2 Summary of descriptive statistics for observed CO_2 emissions in two agroclimatic regions (i.e., Karaj and Debrecen) (02.08.2018–13.09.2018)

Statistic	All data[a]			Debrecen			Karaj		
	CO_2-Obs	Δ	θ	CO_2-Obs	Δ	θ	CO_2-Obs	Δ	θ
Cases	72	72	72	36	36	36	36	36	36
Minimum	53.00	15.00	8.00	53.00	20.00	12.00	138.00	15.00	8.00
Maximum	906.00	26.00	31.00	361.00	26.00	27.00	906.00	26.00	31.00
Range	853.00	11.00	23.00	308.00	6.00	15.00	768.00	11.00	23.00
Median	226.50	20.00	20.00	117.50	24.00	23.00	503.00	19.00	16.50
Mean	308.86	21.24	19.56	138.78	22.83	21.33	478.94	19.64	17.78
Standard deviation (n)	216.10	3.14	5.02	72.04	2.48	4.38	174.22	2.91	4.98
Skewness (Pearson)	0.67	− 0.01	− 0.02	1.45	− 0.14	− 0.67	0.02	0.52	0.65
Kurtosis (Pearson)	− 0.71	− 1.13	− 0.60	1.85	− 1.73	− 0.38	− 0.51	− 0.57	0.57
Standard error of the mean	25.65	0.37	0.60	12.18	0.42	0.74	29.45	0.49	0.84

[a]Δ: soil temperature, θ: soil moisture

Table 3 Scenarios used for predicting CO_2 emissions (CO_2-Prd) in all algorithms

Scenarios/input	Abbreviation	Covariates[a]	Output
First scenario	SC1	SM + Δ + θ	CO_2-Prd
Second scenario	SC2	SM + Δ	CO_2-Prd
Third scenario	SC3	SM + θ	CO_2-Prd

[a]SM: soil management techniques, Δ: soil temperature, θ: soil moisture, CO_2-Prd: predicted values of CO_2

Table 4 Hyperparameters selected in machine and deep learning architectures

Algorithm	Hyperparameters
GBR	n_estimators: [50, 100,] learning_rate: [0.01, 0.1,] max_depth: [3, 5,] min_samples_split: [2, 5, 10] min_samples_leaf: [1, 2, 4] Iterations = 10, cv = 5
SVR	Regularization C: reciprocal (1e−2, 1e2) gamma: expon(scale = 1.0) kernel: ['linear','rbf',] Iterations = 10, cv = 5
FNN	Learning rate = '0.001', neurons = '128', activation = 'relu', Optimizer = 'adam', epochs = 50, batch_size = 32,
CNN	Learning rate = '0.001', filters = '128', kernel size = '3', activation = 'relu', Optimizer = 'adam', epochs = 50, batch_size = 32,

2.2.3 Feedforward Neural Network (FNN)

Feedforward architecture of ANN consists of three layers: [input layer, hidden layer (s), and output layer], where each layer has several neurons that are characterized by different biases and weights (Widiasari and Nugroho 2017). The data flows from the input layer to the output one through the hidden layer (s), process also called as forward pass (Harun et al. 2010). Currently, FNN for CO_2 prediction is employed using the Keras library with a TensorFlow backend in python. An FNN is a foundational architecture in deep learning, comprising layers of interconnected nodes, where information flows unidirectionally from the input layer through hidden layers to the output layer (Khadem et al. 2022).

$$Y_j = (f) * \sum_{i=1}^{n} (X_i . W_{ij}) + b_j, \tag{1}$$

where Y_j is the weighted sum of inputs for neuron j, X_i is the ith neuron in previous layer, W_{ij} is the weight associated between the neurons, and b_j is the introduced bias term in jth neuron, f denotes activation function. The FNN architecture of the current study is defined as a sequential model with an input layer, two hidden layers employing rectified linear unit (ReLU) activation functions, and an output layer. Features are standardized using the scikit-learn StandardScaler. The Adam optimizer is employed with a specified learning rate, and the mean squared error is chosen as the loss function. The model is then trained on standardized training data for 50 epochs, with a batch size of 32 (Table 4). The code also allowed for experimenting with different learning rates, providing flexibility for hyperparameter tuning to optimize the neural network's performance.

2.2.4 Convolutional Neural Network (CNN)

Convolutional neural network is a kind of deep and feedforward artificial neural network that is particularly designed for image processing, classification, and object detection. The basic architecture of CNN comprises of convolutional layer to apply convolutional operation to the input data, and pooling layer to downscale the spatial dimensions (Cui and Fearn 2018). Convolutional layers use filters or kernels for feature extraction and capturing spatial hierarchies. The down sampling process by pooling layer helps to improve the computational complexity. Fully connected layers, traditionally located at the network's end, establish connections between each neuron, and every neuron present in the previous and subsequent layers, thereby facilitating high-level reasoning (Manaswi 2018). During forward propagation, the input of each feature signal in layer l is the accumulation of the final output of the previous feature signal ($l-1$) convolved with proper filters and passed through a nonlinear activation function (Malek et al. 2018). Feedforward architecture of 1D CNN is presented as

$$a_i^l = b_i^l + \sum_{j=1}^{m^{l-1}} conv1D(W_{i,j}^l S_j^{l-1})\left(i = 1 \ldots, m^l\right), \quad (2)$$

where a_i^l presents ith feature signal input, b_i^l is the bias term, $W_{i,j}^l$ is the kernel (filter) weights, S_j^{l-1} is the jth feature output on previous ($l-1$) layer. The currently employed CNN architecture is formed by layers, each having a distinct function in the learning process. The input layer has been customized to accommodate the shape of our preprocessed data. A 1D convolutional layer that consists of 128 filters and a kernel size of 3 is employed which is particularly responsible for learning local patterns and features within the sequential data. The rectified linear unit (ReLU) activation function has been utilized to introduce non-linearity. After the convolutional layer, a flattening layer was introduced to alter the multi-dimensional output into a one-dimensional array. The architecture applied incorporated two dense (fully connected) layers. The first dense layer consisted of 128 units with a ReLU activation function, thereby facilitating the extraction of high-level features. The final dense layer, which consisted of a single unit, served as the output layer for CO_2 prediction. Like FNN, the Adam optimizer is employed with a specified learning rate (0.01), with mean squared error as the loss function. The model was trained for 50 epochs with a batch size of 32.

The input layer included multi factors including soil management techniques (SM), and meteorological factors includes soil temperature (Δ), and soil moisture (θ) with predicted CO_2 as an output layer. (Table 3).

2.2.5 Machine Learning Models Performance

All models' evaluations were performed using four indicators, namely: (1) coefficient of determination (R^2), Root mean squared error (RMSE), degree of agreement (d), and the Nash–Sutcliffe efficiency (NSE) presented in Table 5. Moreover, Taylor diagram (Taylor 2001) was implemented to exhibit the performance of all models in different scenarios of CO_2 predictions. It provides a concrete overview of the model's performances based on the correlation and standard deviation between observed and predicted values (Zhu et al. 2020).

3 Results

3.1 Efficiency of Machine and Deep Learning Models in Predicting CO_2 Emissions

3.1.1 Predicting CO_2 Emissions at Training Stage

Three scenarios were considered to reach the optimal combination between input data and predicted CO_2 emissions. For the first scenario (SC1), all variables (SM + Δ + θ) were treated as input for both machine and deep learning algorithms. Performance evaluation results of train set revealed that high prediction accuracy is achieved by GBR in SC1 (SM + Δ + θ) with the highest $R^2 = 0.8778$, RMSE = 72.05, d = 0.96, and NSE = 0.870 followed by GBR in SC3 (SM + θ) with the $R^2 = 0.854$, RMSE = 78.73, d = 0.950, and NSE = 0.850 (Figs. 2, 3).

Following the sequence of training accuracy, FNN outperformed in SC1 (SM + Δ + θ) with the $R^2 = 0.791$, RMSE = 94.08, d = 0.93, and NSE = 0.790, and in SC3 (SM + θ) $R^2 = 0.757$, RMSE = 101.43, d = 0.920, NSE = 0.750. Furthermore, SVR is found to perform least with $R^2 = 0.660$, RMSE = 120.06, d = 0.860, NSE = 0.660 in SC1 (SM + Δ + θ), and $R^2 = 0.735$, RMSE = 105.9, d = 0.900, NSE = 0.730 in SC3 (SM + θ). SC2 (SM + Δ) provided with least accurate CO_2 predictions with $R^2 = 0.715$ for GBR, followed by $R^2 = 0.653$ for FNN, $R^2 = 0.642$ for CNN, and $R^2 = 0.530$ for SVR (Table 6**,** Figs. 2, 4).

Overall, the performance of machine and deep learning algorithms for high prediction accuracy also presented in Tylor diagram (Fig. 4) is ranked as GBR > FNN > CNN > SVR.

3.1.2 Predicting CO_2 Emissions at Testing Stage

In testing stage, CO_2 predictions obtained on the unseen test data revealed the highest prediction accuracy is achieved by FNN in SC1 (SM + Δ + θ) with $R^2 = 0.918$, RMSE = 67.75, d = 0.97, NSE = 0.910 followed by FNN in SC3 (SM + θ) with $R^2 = 0.907$, RMSE = 72.22, d = 0.97, and NSE = 0.90, and GBR in SC1 (SM + Δ + θ) with $R^2 = 0.887$, RMSE = 79.881, d = 0.960, and NSE = 0.880 (Table 6**,** Fig. 5).

Afterwards, based on performance evaluation, CNN in SC1 (SM + Δ + θ) performed with $R^2 = 0.886$, RMSE = 80.11, d = 0.96, NSE = 0.880, and GBR in SC3

(SM + θ) with $R^2 = 0.859$, RMSE = 88.96, d = 0.950, and NSE = 0.860. Like training stage, SVR is found to perform less compared to other algorithms with the highest achieved $R^2 = 0.809$, RMSE = 103.8, d = 0.92, NSE = 0.80 in the SC3 (SM + θ) followed by $R^2 = 0.690$, RMSE = 132.2, d = 0.860, NSE = 0.690 in SC1(SM + Δ + θ), and $R^2 = 0.515$, RMSE = 165.3 d = 0.760, NSE = 0.510 in SC2 (SM + Δ) (Table 6**,** Figs. 5, 6).

During the prediction of CO_2 emissions, machine learning tends to either overpredict or underpredict emissions compared with the original data. To capture the error, the observed data were compared with the predicted data, and the differences were displayed as a box plot, as seen in Fig. 7. In the training stage, the smallest differences were captured in GBR-SC1 and GBR-SC3, while some cases of overprediction were recorded in CNN-SC3 and NN-MLP-SC2 (Fig. 7A). In the testing stage, the analysis of errors revealed that the smallest errors in predicting CO_2 emissions were observed in FNN-SC1 and GRB-SC2 (Fig. 7B), while overprediction of some values was noted in SVR. In this context, a ridgeline chart confirms this outcome, comparing the distribution of actual data (black line) with other prediction data in each scenario to reveal some differences (Fig. 8). In the training stage, GBR (red color) was found to have the closest distribution to the original. However, SVR exhibited a less similar distribution compared with the original. In the testing stage, FNN (blue) and GRB (red) had the closest distribution compared with the actual data.

Overall, based on the performance evaluation for the prediction accuracy, algorithms are ranked as FNN > GRB > CNN > SVR as also presented in the Tylor diagram (Fig. 4). Moreover, three different scenarios provided a better exploration for identifying the best input combinations. It revealed that SC1 with (SM + Δ + θ) provided the best input combination for achieving high prediction accuracy during model training with optimized hyperparameters and five-fold CV followed by SC3 (SM + θ) and SC2 (SM + Δ). Scatterplot of observed and predicted CO_2 emissions from the four applied algorithms also presents more closeness towards fit line by FNN and GBR. Furthermore, box plots distribution also revealed that FNN provided a more similar range of data distribution with the actual ones followed by GBR and CNN specifically in SC1 and SC3 (Fig. 5). The low performance of SVR is also clearly depicted in the three scenarios. Overall, both deep learning architectures (FNN and CNN) proved a better performance for accurate CO2 predictions compared to SVR.

4 Discussion

4.1 Comparative Performance of Machine and Deep Learning Models for CO_2 Prediction

Currently, implementation of machine and deep learning algorithms for CO_2 emissions revealed the outperformance of GBR in train and FNN in test datasets. The outperformance of GBR for accurate predictions is already proven in various studies in unseen test data (Adjuik and Davis 2022; Romeiko et al. 2019; Wang et al. 2024). The sequential learning process of GBR enables a gradual reduction of errors, with a focus on areas of poor model performance. The repeated process enables the model to capture the hidden non-linear relationship unless highly accurate prediction is achieved. The outperformance of GBR in SC1 (SM + Δ + θ) with highest $R^2 = 0.877$ on train and 0.887 on test data (Figs. 3, 6) aligned with the findings of Kaur Dhaliwal et al. (2022)

Table 5 Performance evaluation metrics for all models for predicting CO_2 emissions

Index	Mathematical equation[a]	Range	Ratting
R^2	$1-\frac{\sum_{i=1}^{n}\left(CO_{2pred}-\overline{CO_{2obs}}\right)^2}{\sum_{i=1}^{n}\left(CO_{2obs}-CO_{2obs}\right)^2}$	0 to + 1	1 is the optimal value
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(CO_{2Obs}-\widehat{CO_{2}}_{Pred}})^2$	0 to ∞	0 is the optimal value
d	$1-\frac{\sum_{i=1}^{n}\left(CO_{2Obs}-CO_{2Prd,t}\right)^2}{\sum_{i=1}^{n}\left(\left\vert CO_{2Pred,}-\overline{CO_{2Obs}}\right\vert+\left\vert CO_{2Obs,}-\overline{CO_{2Obs}}\right\vert\right)^2}$	0–1	1 is the optimal value
NSE	$1-\frac{\sum_{t=1}^{n}\left(CO_{2Prd,t}-CO_{2Obs,t}\right)^2}{\sum_{t=1}^{n}\left(CO_{2Obs,t}-\overline{CO_{2Obs}}\right)^2}$	− ∞ and 1	Very good (0.75 < NSE ≤ 1.00) Good (0.65 < NSE ≤ 0.75) Acceptable (0.50 < NSE ≤ 0.65) Unacceptable (0.40 < NSE ≤ 0.50, or NSE ≤ 0.4)

[a]CO_2-Obs: measured CO_2 in the field, CO_2-Prd: predicting CO_2 emissions; $\overline{CO_{2Prd}}$: mean of CO_2 emissions

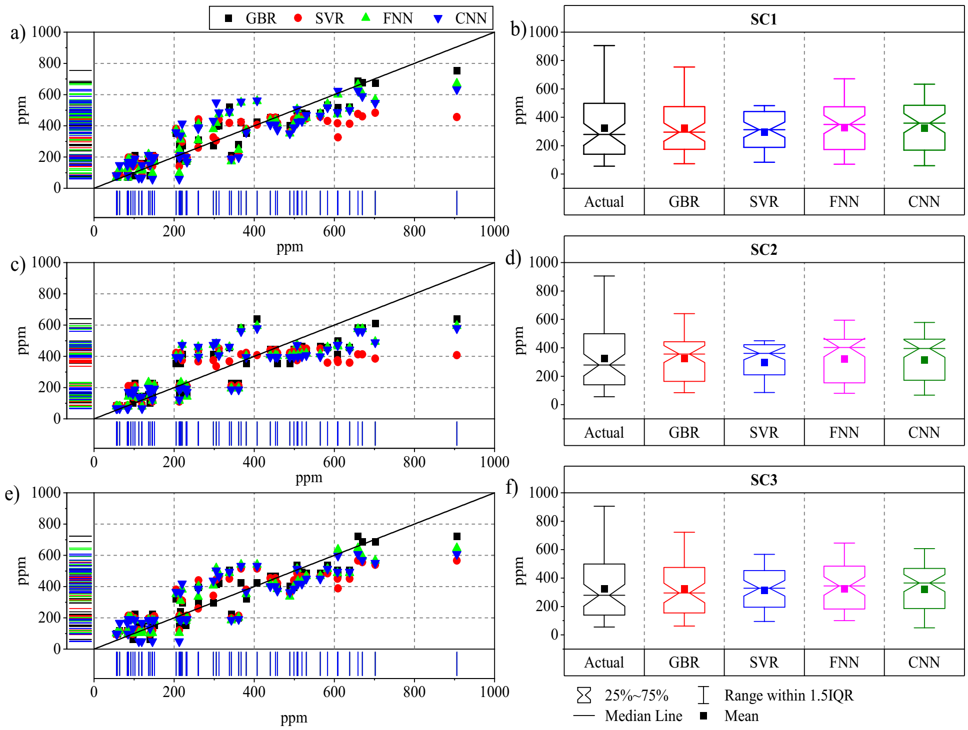

Fig. 2 Scatterplots and box plots distribution of observed (actual) and predicted CO_2 (ppm) emissions in all machine and deep learning models in train set. **a** scatterplot of actual and predicted CO_2 (ppm) emissions from all models in SC1 (SM + Δ + θ). **b** Boxplots showing the mean and range of actual and predicted CO_2 (ppm) emissions from all models in SC1 (SM + Δ + θ). **c** Scatterplot of actual and predicted CO_2 (ppm) emissions from all models in SC2 (SM + Δ). **d** Boxplots showing the mean and range of actual and predicted CO_2 (ppm) emissions from all models in SC2 (SM + Δ). **e** scatterplot of actual and predicted CO_2 (ppm) emissions from all models in SC3 (SM + θ). **f** Boxplots showing the mean and range of actual and predicted CO_2 (ppm) emissions from all models in SC3 (SM + θ)

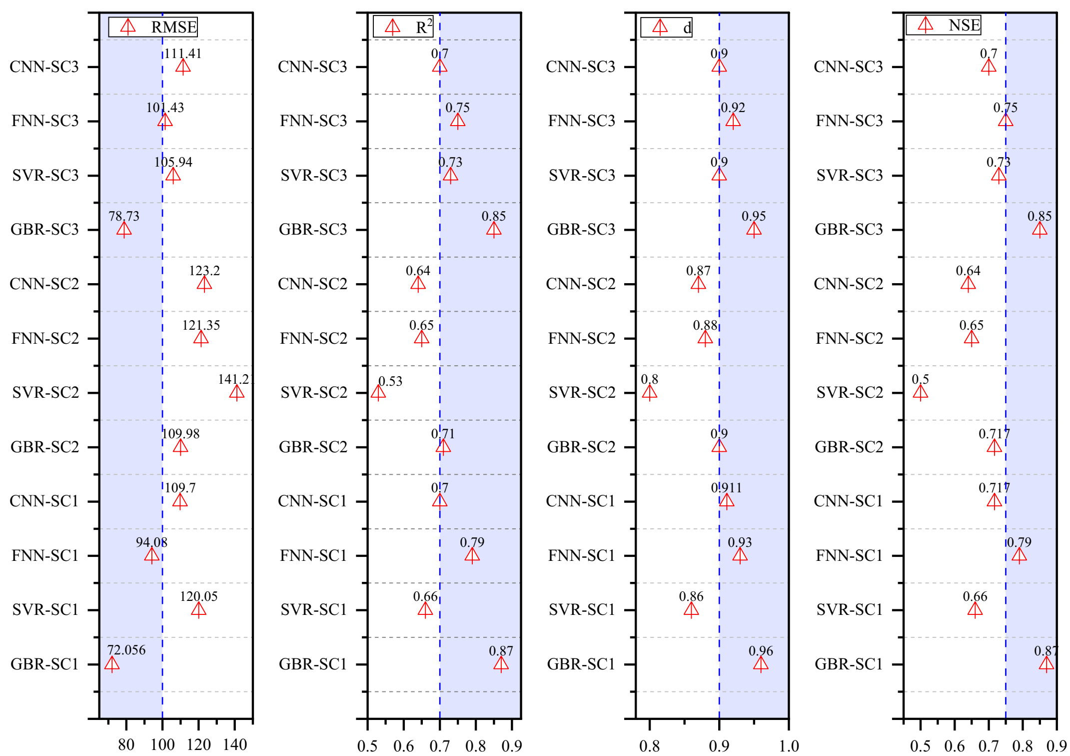

Fig. 3 Performance evaluation metrics (RMSE, R^2, d, and NSE) of all machine and deep learning model in train set

for accurate crop yield prediction from soil management practices. The high performance might be attributed to less sensitivity of GBR towards outliers. This allowed minimizing their influence for accurate CO_2 predictions and helped achieve high accuracy in model training as explained by Ibrahim (2023). Another reason for high

Table 6 Performance evaluation of machine and deep learning algorithms in predicting CO_2 emission in train and test sets

Algorithm	Training				Testing			
	R^2	RMSE	d	NSE	R^2	RMSE	d	NSE
GBR-SC1	0.877	72.05	0.960	0.870	0.887	79.881	0.960	0.880
SVR-SC1	0.660	120.06	0.860	0.660	0.690	132.2	0.860	0.690
FNN-SC1	0.791	94.082	0.930	0.790	0.918	67.756	0.970	0.910
CNN-SC1	0.716	109.71	0.911	0.717	0.886	80.111	0.960	0.880
GBR-SC2	0.715	109.98	0.900	0.717	0.793	108.02	0.920	0.790
SVR-SC2	0.530	141.21	0.800	0.500	0.515	165.39	0.760	0.510
FNN-SC2	0.653	121.36	0.880	0.650	0.787	109.53	0.920	0.780
CNN-SC2	0.642	123.28	0.870	0.640	0.781	110.96	0.910	0.780
GBR-SC3	0.854	78.738	0.950	0.850	0.859	88.961	0.950	0.860
SVR-SC3	0.735	105.94	0.900	0.730	0.809	103.86	0.920	0.800
FNN-SC3	0.757	101.43	0.920	0.750	0.907	72.225	0.970	0.900
CNN-SC3	0.707	111.41	0.900	0.700	0.849	92.125	0.950	0.850

prediction accuracy is automatic feature selection during model training that helped in prioritizing best features with minimized prediction errors (Otchere et al. 2022). Furthermore, mixed data handling ability of GBR also explored by Cha et al. (2021) without extensive preprocessing also lead to capture the best relationship between soil management practices and CO_2 emissions. Support vector regression (SVR) differs from classical regression models as it not only aims to minimize prediction errors but also incorporates a margin of tolerance. This margin allows for deviations from the hyperplane within a specified epsilon-insensitive tube. This inherent flexibility empowers SVR to capture of the overall trend of the data while accommodating minor fluctuations or outliers (Cai and Ma 2019). The effectiveness of SVR lies in its ability to handle non-linear relationships through the utilization of kernel functions, which implicitly map the input features onto a higher-dimensional space where linear separation becomes feasible (Wang and Bi 2023). This procedural approach empowers SVR to comprehend intricate patterns and correlations within the data without explicitly defining the transformation. For instance, Singh and Kumar (2022) also reported the superior performance of SVR with polynomial and RBF kernels for predicting agricultural GHG emissions. However, currently, SVR is observed to perform with a minimum $R^2=0.690$ in best performed SC1 (Figs. 3, 6). The performance of SVR is highly dependent on the appropriate selection of hyperparameters, including the choice of kernel function, regularization parameter (C), and epsilon (Yan et al. 2019).

Furthermore, neural networks (NN) are applied for a comparative assessment and its wide applicability due to their inherent capability of designing qualitative input–output (Küçüktopcu and Cemek 2021). Increasing the depth of NNs and effectively tuning the network may contribute to better prediction performance (Chowdhury et al. 2021). For instance, our findings demonstrated the outperformance of FNN over all algorithms in SC1 ($SM+\Delta+\theta$) with highest $R^2=0.918$, $d=0.97$, and $NSE=0.91$ (Figs. 5, 6). Several distinct capabilities of FNN including scalability, approximating complex functions, and automatic feature learning make them a better choice over a range of classical ML algorithms (Kosamkar and Kulkarni 2021a). Hence, FNNs are more flexible to capture the temporal dependencies and intricate patterns of the data with limited need of manual feature engineering (LeCun et al. 2015). Moreover, parallel processing capabilities due to GPU hardware architecture accelerates the optimization process and reduces the model training timings. Currently used FNN architecture of two hidden layers with 128 neurons each proved to better optimize the model performance achieving highest prediction accuracy.

The unique architecture of CNN has revolutionized image processing due to its exceptional capability of capturing local patterns through filters. Currently, 1D convolutions employed in temporal dimensions of CO_2 were able to extract the local patterns within the time series, identifying the intricate relationships between the soil management practices and CO_2 emissions. Overall, the performance of CNN is ranked 3rd for accurate predictions of CO_2 emissions on test dataset with $R^2=0.886$ in best performed SC1 with a close competition to GBR with $R^2=0.887$ (Figs. 3, 5, 6). The current high performance of deep learning algorithms is aligned with the previous studies of Altikat (2021) and Küçüktopcu and Cemek (2021) for CO_2 predictions in different environmental settings. Specifically, the high performance of CNN is also attributed to its ability of hierarchical representation learning, where lower layers capture simple temporal patterns and deep layers learn more complex relationships (Jing et al. 2017). Moreover, CNNs do not apply parameter sharing in convolutional filters to decrease the number of trainable parameters when compared to fully

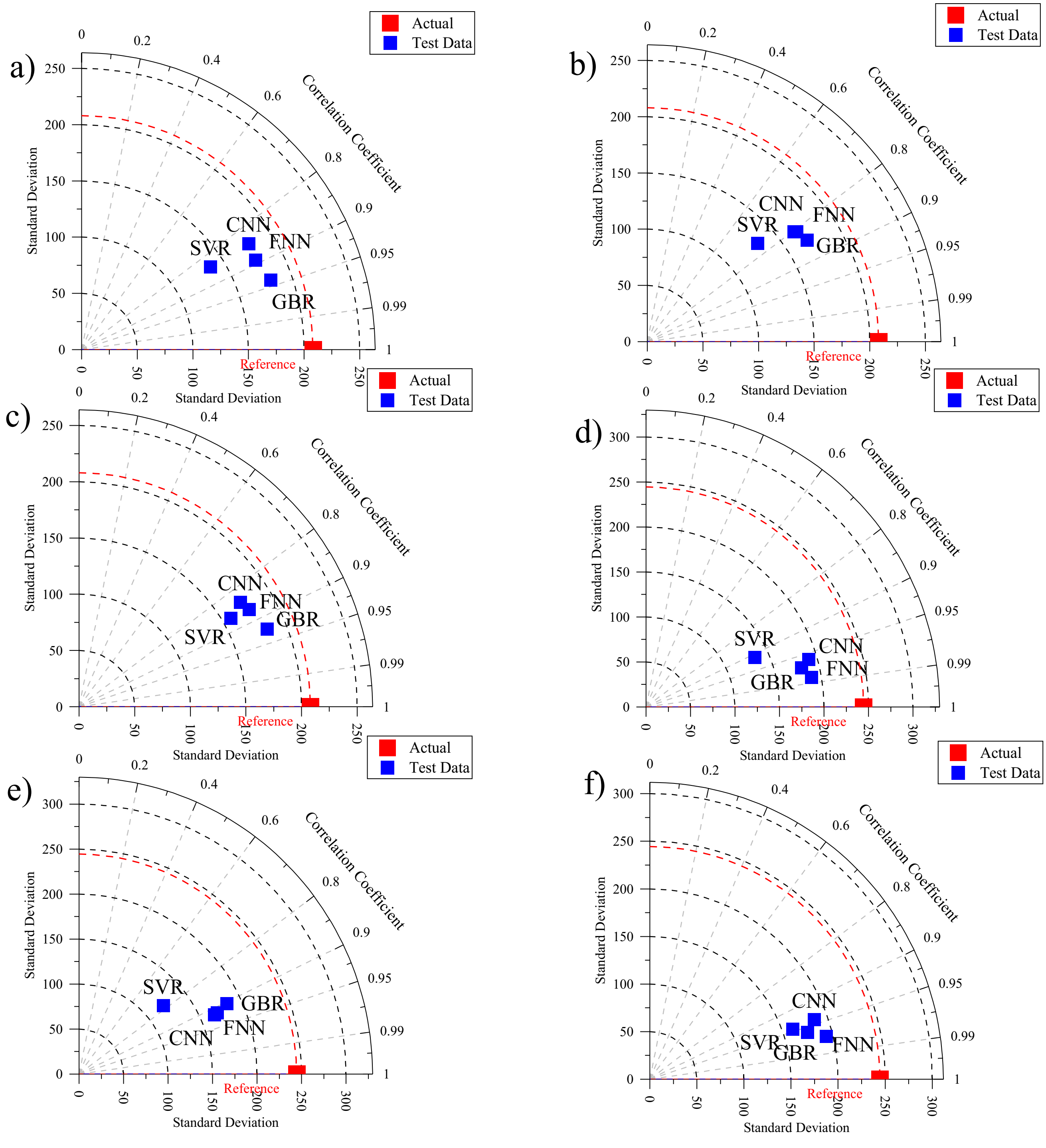

Fig. 4 Tylor diagram of machine learning performance: **a** SC1 training stage, **b** SC2 training stage, **c** SC3 training stage; **d** SC1 testing stage, **e** SC2 testing stage, **f** SC3 testing stage

connected networks. This efficiency in parameters makes CNNs a good option for modeling time series data, particularly with large datasets. However, the limited global context of the model may reduce its capability to capture long range dependencies in the time series affecting their predictive performance. Hence, in some CO_2 predictions studies, hybrid of CNN-LSTM is reported to perform better (Amarpuri et al. 2019; Han et al. 2023). Overall, FNN has proven to be the most robust and effective method for accurate CO_2 prediction on unseen test dataset attributed to it high flexibility towards capturing complex relationship. Hence, accurate prediction modelling using advanced

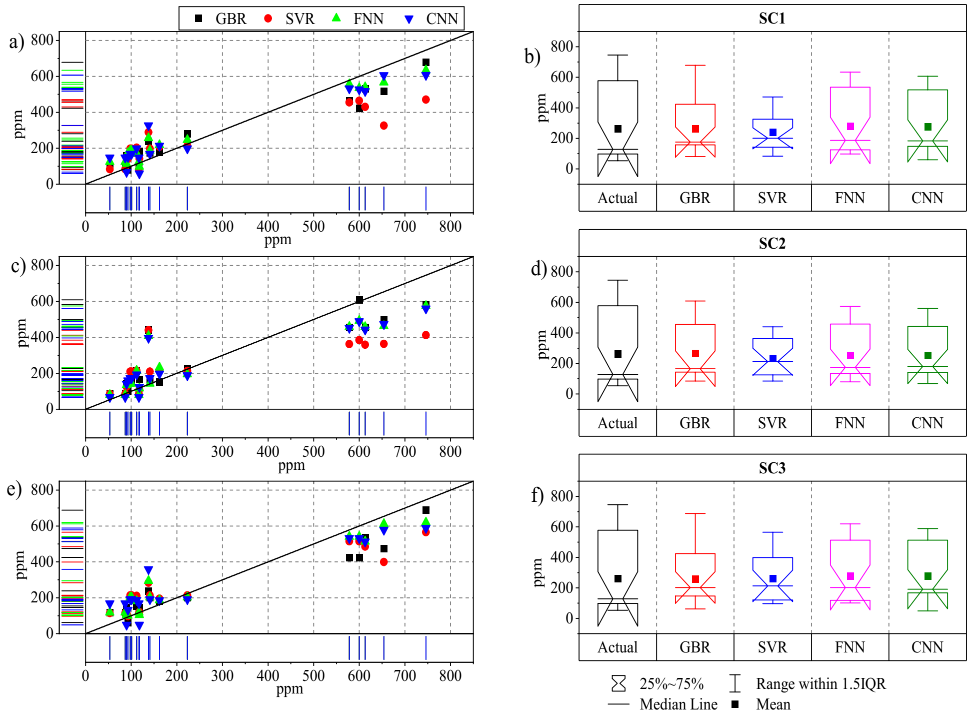

Fig. 5 Scatterplots and box plots distribution of observed (actual) and predicted CO_2 (ppm) emissions in all machine and deep learning models in test set. **a** scatterplot of actual and predicted CO_2 (ppm) emissions from all models in SC1 (SM + Δ + θ). **b** Boxplots showing the mean and range of actual and predicted CO_2 (ppm) emissions from all models in SC1 (SM + Δ + θ). **c** scatterplot of actual and predicted CO_2 (ppm) emissions from all models in SC2 (SM + Δ). **d** Boxplots showing the mean and range of actual and predicted CO_2 (ppm) emissions from all models in SC2 (SM + Δ). **e** scatterplot of actual and predicted CO_2 (ppm) emissions from all models in SC3 (SM + θ). **f** Boxplots showing the mean and range of actual and predicted CO_2 (ppm) emissions from all models in SC3 (SM + θ)

machine and deep learning approaches can facilitate for an informed decision-making process for adopting effective soil management practices. Moreover, despite good hyperparameter adjustments, each of the algorithms is bounded with some limitations addressed further in limitations.

4.2 Limitations

In this research, only soil parameters were considered, specifically soil moisture and soil temperature. These data were collected in the field during measurements of CO_2 emissions. Other factors, such as climatic variables (temperature, relative humidity, and precipitation), were not considered due to the absence of a climatic monitoring station in the field. Incorporating these factors may enhance the performance of the studied algorithms and highlight the significant importance of these variables. Therefore, future studies will focus on the role of climatic factors in emissions.

Despite significant high predictive performance of GBR and FNN in train and test sets for CO_2 emissions, the algorithms are bound to some limitations. Hyperparameters sensitivity of GBR is one of the main challenge when dealing with complex relationships or large datasets which may also cause overfitting (Malinin et al. 2020). Similarly, in contrast to the advantageous capabilities of deeply learned NNs, high dimensional data structures also cause model overfitting issues. Manual hyperparameter tuning including no. of layers and neurons, learning rates, and regularization parameters can also enhance the sensitivity towards prediction accuracies (Ying 2019). Due to black box nature of machine and deep learning models, interpretability of the features is challenging, especially in complex architectures with many hidden layers (Gawlikowski et al. 2021). Hence, the addition of more covariates facilitated by feature selection method can be used for better interpretation of models.

Another significant factor affecting the prediction accuracies of deep learning models is the length of dataset which can be facilitated by data augmentation methods to prevent overfitting problems. Furthermore, SVR might present computational hurdles, particularly for extensive datasets, owing to the quadratic optimization issue it deals with (Awad and Khanna 2015). Overall, deep learning provided reliable and accurate findings with better performance of FNN which can be further improved by increasing the sampled observations. Furthermore, incorporation of soil physiochemical and biological properties can also increase the prediction accuracy of CO_2 emission from agricultural soil (Ebrahimi et al. 2019).

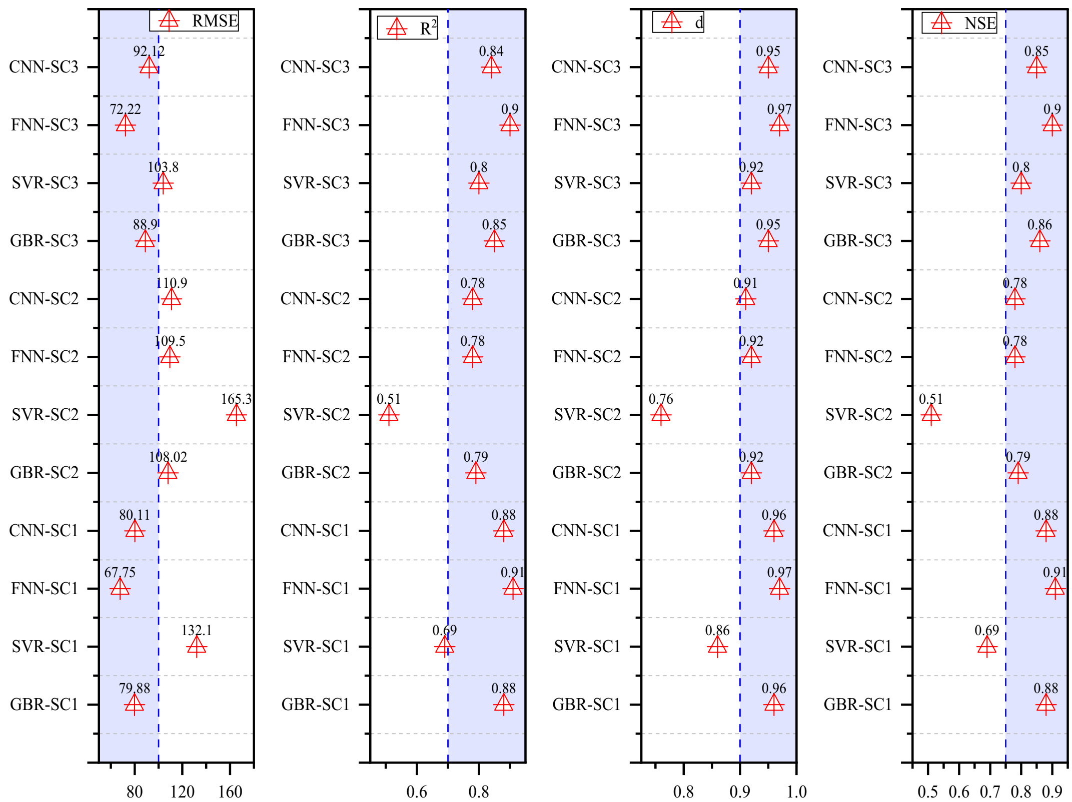

Fig. 6 Performance evaluation metrics (RMSE, R^2, d, and NSE) of all machine and deep learning model in test set

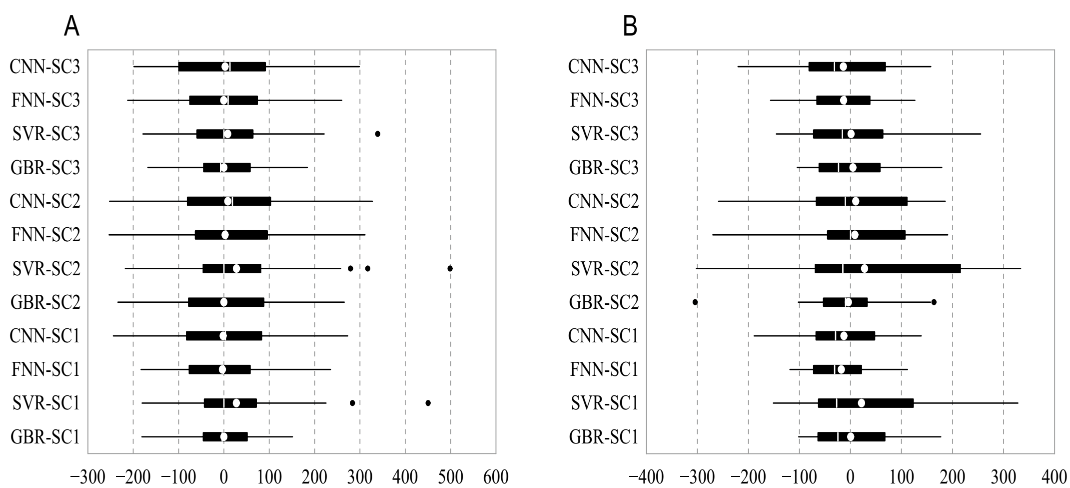

Fig. 7 Box plot displaying the errors in prediction compared with observed data for different scenarios: **A** training, **B** testing

5 Conclusion

Recently, the use of machine learning (ML) in environmental research has accelerated. This research utilized two machine learning models, namely GBR and support vector regression (SVR), along with two deep learning architectures, FNN and CNN, to predict CO_2 emissions from two agroclimatic regions. The results revealed that FNN outperformed the others, achieving the highest accuracy in predicting CO_2 emissions, followed by GBR and CNN. SVR's performance was the least accurate. Additionally, this research recommended using a combination, SC1 ($SM + \Delta + \theta$), which was found to provide highly accurate predictions in both

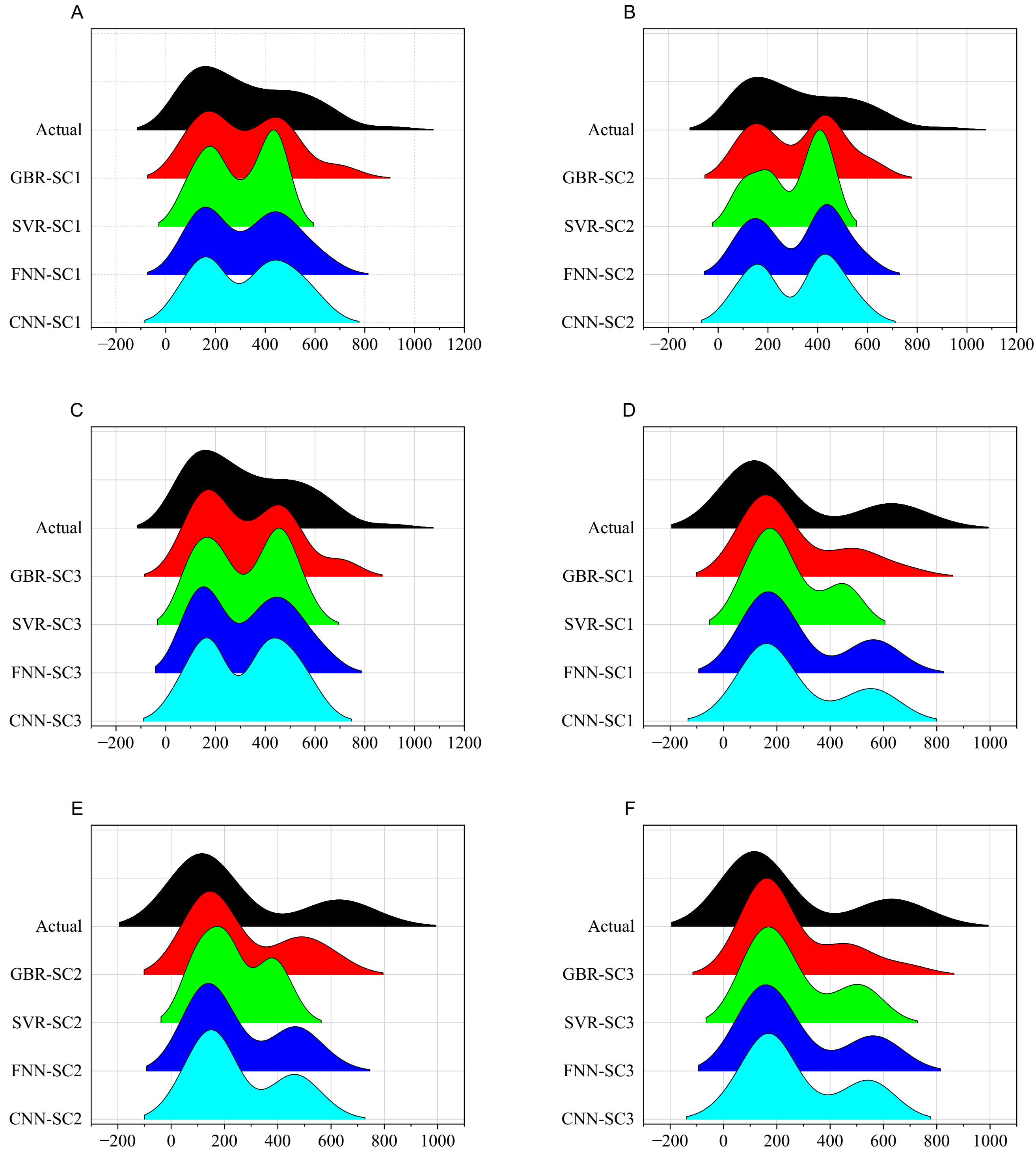

Fig. 8 Ridgeline chart comparing the distribution of actual data with machine learning predicted data for different scenarios: **A** SC1 (SM + Δ + θ, training), **B** SC2 (SM + Δ, training), **C** SC3 (SM + θ, training), **D** SC1 (SM + Δ + θ, testing), **E** SC2 (SM + Δ, testing), **F** SC3 (SM + θ, testing)

training and testing datasets, reflecting the significance of all factors—soil management and meteorological—in predicting CO_2 emissions across different environmental settings. Interestingly, the input combination of soil management practices with soil temperature yielded less accurate predictions, highlighting the more significant impact of soil moisture on CO_2 emissions. Predicting CO_2 emissions from agricultural soil from varied agroclimatic regions poses significant implications for agricultural sustainability and addressing climate change. The outcome of this research will serve as a tool to support decision-makers, researchers, and stakeholders in assessing the capability of ML for accurate CO_2 prediction. Predictive modeling that identifies the most influential factors affecting CO_2 emissions can assist in developing strategies to increase soil carbon storage and decrease greenhouse gas emissions from agriculture.

Acknowledgements Authors from the University of Debrecen would like to thank the University of Debrecen Program (DETKA) for the Scientific Research Bridging Fund.

Funding Open access funding provided by University of Debrecen. Covered for reviewing process. Project no. TKP2021-NKTA-32 has been implemented with support from Hungary's National Research, Development, and Innovation Fund, financed under the TKP2021-NKTA funding scheme.

Data availability Data available upon request from the correspondence authors.

Declarations

Conflict of interest The author(s) 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

Abbasi NA, Hamrani A, Madramootoo CA, Zhang T, Tan CS, Goyal MK (2021) Modelling carbon dioxide emissions under a maize-soy rotation using machine learning. Biosyst Eng 212:1–18. https://doi.org/10.1016/j.biosystemseng.2021.09.013

Adjuik TA, Davis SC (2022) Machine learning approach to simulate soil CO_2 fluxes under cropping systems. Agronomy 12:197

Almagro M, Querejeta JI, Boix-Fayos C, Martínez-Mena M (2013) Links between vegetation patterns, soil C and N pools and respiration rate under three different land uses in a dry Mediterranean ecosystem. J Soils Sediments 13:641–653

Altikat S (2021) Prediction of CO_2 emission from greenhouse to atmosphere with artificial neural networks and deep learning neural networks. Int J Environ Sci Technol 18:3169–3178. https://doi.org/10.1007/s13762-020-03079-z

Amarpuri L, Yadav N, Kumar G, Agrawal S (2019) Prediction of CO_2 emissions using deep learning hybrid approach: a case study in Indian context. In: 2019 twelfth international conference on contemporary computing (IC3), pp 1–6

Arshad S, Kazmi JH, Javed MG, Mohammed S (2023a) Applicability of machine learning techniques in predicting wheat yield based on remote sensing and climate data in Pakistan, South Asia. Eur J Agron 147:126837. https://doi.org/10.1016/j.eja.2023.126837

Arshad S, Kazmi JH, Prodhan FA, Mohammed S (2023b) Exploring dynamic response of agrometeorological droughts towards winter wheat yield loss risk using machine learning approach at a regional scale in Pakistan. Field Crop Res 302:109057. https://doi.org/10.1016/j.fcr.2023.109057

Awad M, Khanna R (2015) Support vector regression. In: Awad M, Khanna R (eds) Efficient learning machines: theories, concepts, and applications for engineers and system designers. Apress, Berkeley, pp 67–80

Bååth E (2018) Temperature sensitivity of soil microbial activity modeled by the square root equation as a unifying model to differentiate between direct temperature effects and microbial community adaptation. Glob Change Biol 24:2850–2861. https://doi.org/10.1111/gcb.14285

Bhattacharyya SS, Leite FFGD, France CL, Adekoya AO, Ros GH, de Vries W, Melchor-Martínez EM, Iqbal HMN, Parra-Saldívar R (2022) Soil carbon sequestration, greenhouse gas emissions, and water pollution under different tillage practices. Sci Total Environ 826:154161. https://doi.org/10.1016/j.scitotenv.2022.154161

Cai J, Ma X (2019) Carbon emission prediction model of agroforestry ecosystem based on support vector regression machine. Appl Ecol Environ Res. https://doi.org/10.15666/AEER/1703_63976413

Campbell B, Chen L, Dygert C, Dick W (2014) Tillage and crop rotation impacts on greenhouse gas fluxes from soil at two long-term agronomic experimental sites in Ohio. J Soil Water Conserv 69:543–552

Canteral KFF, Vicentini ME, de Lucena WB, de Moraes MLT, Montanari R, Ferraudo AS, Peruzzi NJ, La Scala N, Panosso AR (2023) Machine learning for prediction of soil CO_2 emission in tropical forests in the Brazilian Cerrado. Environ Sci Pollut Res 30:61052–61071. https://doi.org/10.1007/s11356-023-26824-6

Cha G-W, Moon H-J, Kim Y-C (2021) Comparison of random forest and gradient boosting machine models for predicting demolition waste based on small datasets and categorical variables. Int J Environ Res Public Health 18:8530. https://doi.org/10.3390/ijerph18168530

Chowdhury S, Rubi MA, Bijoy MHI (2021) Application of artificial neural network for predicting agricultural methane and CO_2 emissions in Bangladesh. In: 2021 12th international conference on computing communication and networking technologies (ICCCNT), pp 1–5

Cui C, Fearn T (2018) Modern practical convolutional neural networks for multivariate regression: applications to NIR calibration. Chemometr Intell Lab Syst 182:9–20. https://doi.org/10.1016/j.chemolab.2018.07.008

Devkota M, Singh Y, Yigezu YA, Bashour I, Mussadek R, Mrabet R (2022) Chapter five—conservation agriculture in the drylands of the Middle East and North Africa (MENA) region: past trend, current opportunities, challenges and future outlook. In: Sparks DL (ed) Advances in agronomy. Academic Press, New York, pp 253–305

Durandeau S, Gabrielle B, Godard C, Jayet PA, Le Bas C (2010) Coupling biophysical and micro-economic models to assess the effect of mitigation measures on greenhouse gas emissions from agriculture. Clim Change 98:51–73. https://doi.org/10.1007/s10584-009-9653-8

Ebrahimi M, Sarikhani MR, Safari Sinegani AA, Ahmadi A, Keesstra S (2019) Estimating the soil respiration under different land uses using artificial neural network and linear regression models. CATENA 174:371–382. https://doi.org/10.1016/j.catena.2018.11.035

Fageria NK, Moreira A (2011) Chapter four—the role of mineral nutrition on root growth of crop plants. In: Sparks DL (ed) Advances in agronomy. Academic Press, New York, pp 251–331

Fattah M, Morshed SR, Morshed SY (2021) Multi-layer perceptron-Markov chain-based artificial neural network for modelling future land-specific carbon emission pattern and its influences on surface temperature. SN Appl Sci 3:1–22

Freitas LPS, Lopes MLM, Carvalho LB, Panosso AR, La Scala JN, Freitas RLB, Minussi CR, Lotufo ADP (2018) Forecasting the spatiotemporal variability of soil CO2 emissions in sugarcane areas in southeastern Brazil using artificial neural networks. Environ Monit Assess 190:741. https://doi.org/10.1007/s10661-018-7118-0

Ganzenmüller R, Pradhan P, Kropp JP (2019) Sectoral performance analysis of national greenhouse gas emission inventories by means of neural networks. Sci Total Environ 656:80–89. https://doi.org/10.1016/j.scitotenv.2018.11.311

Gawlikowski J, Tassi CRN, Ali M, Lee J, Humt M, Feng J, Kruspe A, Triebel R, Jung P, Roscher R (2021) A survey of uncertainty in deep neural networks. arXiv preprint arXiv:210703342.

Gholami V, Booij M, Tehrani EN, Hadian M (2018) Spatial soil erosion estimation using an artificial neural network (ANN) and field plot data. CATENA 163:210–218

Hamrani A, Akbarzadeh A, Madramootoo CA (2020) Machine learning for predicting greenhouse gas emissions from agricultural soils. Sci Total Environ 741:140338. https://doi.org/10.1016/j.scitotenv.2020.140338

Han Z, Cui B, Xu L, Wang J, Guo Z (2023) Coupling LSTM and CNN neural networks for accurate carbon emission prediction in 30 Chinese provinces. Sustainability 15:13934

Harun N, Woo WL, Dlay S (2010) Performance of keystroke biometrics authentication system using artificial neural network (ANN) and distance classifier method. In: International conference on computer and communication engineering (ICCCE'10). IEEE, pp 1–6

Hu R, Fang F, Pain CC, Navon IM (2019) Rapid spatio-temporal flood prediction and uncertainty quantification using a deep learning method. J Hydrol 575:911–920. https://doi.org/10.1016/j.jhydrol.2019.05.087

Ibrahim AF (2023) Prediction of shale wettability using different machine learning techniques for the application of CO_2 sequestration. Int J Coal Geol 276:104318. https://doi.org/10.1016/j.coal.2023.104318

Iqbal M, Naeem UA, Ahmad A, Ghani U, Farid T (2020) Relating groundwater levels with meteorological parameters using ANN technique. Measurement 166:108163

Jerome HF (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29:1189–1232. https://doi.org/10.1214/aos/1013203451

Jiang Q, Qi Z, Madramootoo CA, Smith W, Abbasi NA, Zhang T-Q (2018) Comparison of RZWQM2 and DNDC model in simulating greenhouse gas emission, crop yield and subsurface drainage. In: 2018 ASABE annual international meeting. American Society of Agricultural and Biological Engineers, p 1

Jing L, Zhao M, Li P, Xu X (2017) A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. Measurement 111:1–10. https://doi.org/10.1016/j.measurement.2017.07.017

Kaur Dhaliwal J, Panday D, Saha D, Lee J, Jagadamma S, Schaeffer S, Mengistu A (2022) Predicting and interpreting cotton yield and its determinants under long-term conservation management practices using machine learning. Comput Electron Agric 199:107107. https://doi.org/10.1016/j.compag.2022.107107

Khadem SA, Bensebaa F, Pelletier N (2022) Optimized feed-forward neural networks to address CO_2-equivalent emissions data gaps—application to emissions prediction for unit processes of fuel life cycle inventories for Canadian provinces. J Clean Prod 332:130053. https://doi.org/10.1016/j.jclepro.2021.130053

Kim M, Gilley JE (2008) Artificial neural network estimation of soil erosion and nutrient concentrations in runoff from land application areas. Comput Electron Agric 64:268–275

Kosamkar PK, Kulkarni VY (2021a) Analysis and prediction of greenhouse gas emission using feedforward neural network. Inf Technol Ind 9:304–313

Kosamkar PK, Kulkarni VY (2021b) Effect of soil and climatic attribute on greenhouse gas emission from agriculture sector. In: Bhateja V, Peng S-L, Satapathy SC, Zhang Y-D (eds) Evolution in computational intelligence. Springer Singapore, Singapore, pp 91–101

Küçüktopcu E, Cemek B (2021) The use of artificial neural networks to estimate optimum insulation thickness, energy savings, and carbon dioxide emissions. Environ Prog Sustain Energy 40:e13478

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539

Li H, Qiu J, Wang L, Tang H, Li C, Van Ranst E (2010a) Modelling impacts of alternative farming management practices on greenhouse gas emissions from a winter wheat–maize rotation system in China. Agric Ecosyst Environ 135:24–33. https://doi.org/10.1016/j.agee.2009.08.003

Li H, Qiu J, Wang L, Tang H, Li C, Van Ranst E (2010b) Modelling impacts of alternative farming management practices on greenhouse gas emissions from a winter wheat–maize rotation system in China. Agric Ecosyst Environ 135:24–33

Li J, Wang E, Wang Y, Xing H, Wang D, Wang L, Gao C (2016) Reducing greenhouse gas emissions from a wheat–maize rotation system while still maintaining productivity. Agric Syst 145:90–98. https://doi.org/10.1016/j.agsy.2016.03.007

Li H, Zhang Y, Sun Y, Zhang Q, Liu P, Wang X, Li J, Wang R (2022) No-tillage with straw mulching improved grain yield by reducing soil water evaporation in the fallow period: a 12-year study on the Loess Plateau. Soil Tillage Res 224: https://doi.org/10.1016/j.still.2022.105504

Lin X, Hou J, Wu X, Lin D (2024) Elucidating the impacts of microplastics on soil greenhouse gas emissions through automatic machine learning frameworks. Sci Total Environ 916:170308. https://doi.org/10.1016/j.scitotenv.2024.170308

Lloyd K, Madramootoo CA, Edwards KP, Grant A (2019) Greenhouse gas emissions from selected horticultural production systems in a cold temperate climate. Geoderma 349:45–55

Madu CN, Kuei C-h, Lee P (2017) Urban sustainability management: a deep learning perspective. Sustain Cities Soc 30:1–17. https://doi.org/10.1016/j.scs.2016.12.012

Malek S, Melgani F, Bazi Y (2018) One-dimensional convolutional neural networks for spectroscopic signal regression. J Chemom 32:e2977. https://doi.org/10.1002/cem.2977

Malinin A, Prokhorenkova L, Ustimenko A (2020) Uncertainty in gradient boosting via ensembles. arXiv preprint arXiv:200610562. https://doi.org/10.48550/arXiv.2006.10562

Manaswi NK (2018) Convolutional neural networks. In: Manaswi NK (ed) Deep learning with applications using Python: chatbots and face, object, and speech recognition with Tensorflow and Keras. Apress, Berkeley, pp 91–96

Mancinelli R, Marinari S, Atait M, Petroselli V, Chilosi G, Jasarevic M, Catalani A, Abideen Z, Mirzaei M, Allam M (2023) Durum wheat–potato crop rotation, soil tillage, and fertilization source affect soil CO2 emission and C storage in the Mediterranean environment. Land 12:326. https://doi.org/10.3390/land12020326

Mbow H-OP, Reisinger A, Canadell J, O'Brien P (2017) Special Report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems (SR2). Ginevra, IPCC 650

Mirzaei M, Gorji Anari M, Diaz-Pines E, Saronjic N, Mohammed S, Szabo S, Nasir Mousavi SM, Caballero-Calvo A (2023) Assessment of soil CO_2 and NO fluxes in a semi-arid region using machine learning approaches. J Arid Environ 211:104947. https://doi.org/10.1016/j.jaridenv.2023.104947

Mirzai M, Gorji M, Moghiseh E, Asadi H, Razavi Toosi E (2022) Sustainable soil management and its role in mitigating greenhouse emissions. Land Manag J 9:187–220

Mitchell J, Reicosky D, Kueneman E, Fisher J, Beck D (2019) Conservation agriculture systems. CABI Rev. https://doi.org/10.1079/PAVSNNR201914001

Mohammed S, Alsafadi K, Enaruvbe GO, Bashir B, Elbeltagi A, Széles A, Alsalman A, Harsanyi E (2022a) Assessing the impacts of agricultural drought (SPI/SPEI) on maize and wheat yields across Hungary. Sci Rep 12:8838. https://doi.org/10.1038/s41598-022-12799-w

Mohammed S, Mirzaei M, Pappné Törő Á, Anari MG, Moghiseh E, Asadi H, Szabó S, Kakuszi-Széles A, Harsányi E (2022b) Soil carbon dioxide emissions from maize (*Zea mays* L.) fields as influenced by tillage management and climate. Irrig Drain 71:228–240

Mohammed S, Arshad S, Bashir B, Vad A, Alsalman A, Harsányi E (2024) Machine learning driven forecasts of agricultural water quality from rainfall ionic characteristics in Central Europe. Agric Water Manag 293:108690. https://doi.org/10.1016/j.agwat.2024.108690

Naderi A, Delavar MA, Kaboudin B, Askari MS (2017) Assessment of spatial distribution of soil heavy metals using ANN-GA, MSLR and satellite imagery. Environ Monit Assess 189:1–16

Necpálová M, Anex RP, Fienen MN, Del Grosso SJ, Castellano MJ, Sawyer JE, Iqbal J, Pantoja JL, Barker DW (2015) Understanding the DayCent model: calibration, sensitivity, and identifiability through inverse modeling. Environ Model Softw 66:110–130

Otchere DA, Ganat TOA, Ojero 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 Petrol Sci Eng 208:109244. https://doi.org/10.1016/j.petrol.2021.109244

Pradhan P, Tingsanchali T, Shrestha S (2020) Evaluation of soil and water assessment tool and artificial neural network models for hydrologic simulation in different climatic regions of Asia. Sci Total Environ 701:134308. https://doi.org/10.1016/j.scitotenv.2019.134308

Ray RL, Griffin RW, Fares A, Elhassan A, Awal R, Woldesenbet S, Risch E (2020) Soil CO_2 emission in response to organic amendments, temperature, and rainfall. Sci Rep 10:5849. https://doi.org/10.1038/s41598-020-62267-6

Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat (2019) Deep learning and process understanding for data-driven Earth system science. Nature 566:195–204. https://doi.org/10.1038/s41586-019-0912-1

Romeiko XX, Guo Z, Pang Y (2019) Comparison of support vector machine and gradient boosting regression tree for predicting spatially explicit life cycle global warming and eutrophication impacts: a case study in corn production. In: 2019 IEEE international conference on big data (big data), pp 3277–3284

Saha D, Basso B, Robertson GP (2021) Machine learning improves predictions of agricultural nitrous oxide ($_{N2}$O) emissions from intensively managed cropping systems. Environ Res Lett 16:024004. https://doi.org/10.1088/1748-9326/abd2f3

Sainju UM, Jabro JD, Stevens WB (2008) Soil carbon dioxide emission and carbon content as affected by irrigation, tillage, cropping system, and nitrogen fertilization. J Environ Qual 37:98–106

Sauerbeck DR (2001) CO_2 emissions and C sequestration by agriculture – perspectives and limitations. Nutr Cycl Agroecosyst 60:253–266. https://doi.org/10.1023/A:1012617516477

Shabani E, Hayati B, Pishbahar E, Ghorbani MA, Ghahremanzadeh M (2021) A novel approach to predict CO_2 emission in the agriculture sector of Iran based on inclusive multiple model. J Clean Prod 279:123708. https://doi.org/10.1016/j.jclepro.2020.123708

Shakoor A, Shahbaz M, Farooq TH, Sahar NE, Shahzad SM, Altaf MM, Ashraf M (2021) A global meta-analysis of greenhouse gases emission and crop yield under no-tillage as compared to conventional tillage. Sci Total Environ 750:142299. https://doi.org/10.1016/j.scitotenv.2020.142299

Singh AMKJ, Kumar S (2022) Prediction of greenhouse gas emissions from agricultural fields with and without cover crops. Soil Sci Soc Am J 86:1227–1240. https://doi.org/10.1002/saj2.20429

Singh A, Patel S, Bhadani V, Kumar V, Gaurav K (2024) AutoML-GWL: automated machine learning model for the prediction of groundwater level. Eng Appl Artif Intell 127:107405. https://doi.org/10.1016/j.engappai.2023.107405

Smith WN, Grant BB, Desjardins RL, Worth D, Li C, Boles SH, Huffman EC (2010) A tool to link agricultural activity data with the DNDC model to estimate GHG emission factors in Canada. Agric Ecosyst Environ 136:301–309. https://doi.org/10.1016/j.agee.2009.12.008

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmos 106:7183–7192

Törő Á, Ragán P, Rátonyi T, Kith K, Harsányi E (2019) Daily soil carbon dioxide flux under different tillage conditions. Acta Agraria Debreceniensis 2:141–144. https://doi.org/10.34101/actaagrar/2/3693

Tubiello FN, Salvatore M, Rossi S, Ferrara A, Fitton N, Smith P (2013) The FAOSTAT database of greenhouse gas emissions from agriculture. Environ Res Lett 8:015009. https://doi.org/10.1088/1748-9326/8/1/015009

Vitória ELD, Simon CDP, Lacerda EDG, Freitas ILDJ, Gontijo I (2022) Estimating CO_2 emissions from tilled soils through artificial neural networks and multiple linear regression. Revista Caatinga 35:964–973

Wang H, Williams-Stroud S, Crandall D, Chen C (2024) Machine learning and deep learning for mineralogy interpretation and CO_2 saturation estimation in geological carbon Storage: a case study in the Illinois Basin. Fuel 361:130586. https://doi.org/10.1016/j.fuel.2023.130586

Wang X, Bi M (2023) Greenhouse gas prediction method based on particle swarm optimized SVR. ProcSPIE, p 1251126

Wen H-T, Lu J-H, Jhang D-S (2021) Features importance analysis of diesel vehicles' NO_x and CO_2 emission predictions in real road driving based on gradient boosting regression model. Int J Environ Res Public Health 18:13044. https://doi.org/10.3390/ijerph182413044

Widiasari IR, Nugroho LE (2017) Deep learning multilayer perceptron (MLP) for flood prediction model using wireless sensor network based hydrology time series data mining. In: 2017 international conference on innovative and creative information technology (ICITech). IEEE, pp 1–5

Yadav D, Wang J (2017) Modelling carbon dioxide emissions from agricultural soils in Canada. Environ Pollut 230:1040–1049. https://doi.org/10.1016/j.envpol.2017.07.066

Yan C, Shen X, Guo F, Zhao S, Zhang L (2019) A novel model modification method for support vector regression based on radial basis functions. Struct Multidiscip Optim 60:983–997. https://doi.org/10.1007/s00158-019-02251-5

Yan H, Zhang J, Rahman SS, Zhou N, Suo Y (2020) Predicting permeability changes with injecting CO_2 in coal seams during CO_2 geological sequestration: a comparative study among six SVM-based hybrid models. Sci Total Environ 705:135941

Yılmaz G, Bilgili AV (2018) Modeling seasonal variations of long-term soil CO_2 emissions in an orchard plantation in a semiarid area, SE Turkey. Environ Monit Assess 190:1–14

Ying X (2019) An overview of overfitting and its solutions. J Phys Conf Ser 1168:022022. https://doi.org/10.1088/1742-6596/1168/2/022022

Yu K, Hiscox A, DeLaune RD (2013) Greenhouse gas emission by static chamber and eddy flux methods. Methods Biogeochem Wetl 10:427–437

Zaman M, Heng L, Müller C (2021) Measuring emission of agricultural greenhouse gases and developing mitigation options using nuclear and related techniques: applications of nuclear techniques for GHGs. Springer Nature, Cham

Zechmeister-Boltenstern S, Díaz-Pinés E, Spann C, Hofmann K, Schnecker J, Reinsch S (2018) Soil—the hidden part of climate: microbial processes regulating soil–atmosphere exchange of greenhouse gases. Soil and climate. CRC Press, Boca Raton, pp 11–60

Zhang F, O'Donnell LJ (2020) Chapter 7—support vector regression. In: Mechelli A, Vieira S (eds) Machine learning. Academic Press, New York, pp 123–140

Zhou Y, Xia Q, Zhang Z, Quan M, Li H (2022) Artificial intelligence and machine learning for the green development of agriculture in the emerging manufacturing industry in the IoT platform. Acta Agric Scand Sect B Soil Plant Sci 72:284–299. https://doi.org/10.1080/09064710.2021.2008482

Zhu H, Jiang Z, Li J, Li W, Sun C, Li L (2020) Does CMIP6 inspire more confidence in simulating climate extremes over China? Adv Atmos Sci 37:1119–1132

Prof. Dr. Endre Harsányi is a Vice-Rector for Agricultural Innovation and Training Development at the University of Debrecen. He specializes in agricultural sciences and has contributed significantly to the development of precision farming technologies and education in Hungary (https://unideb.hu/en/dr-endre-harsanyi).

Dr. Morad Mirzaei is associated with the School of Natural Sciences at Trinity College Dublin and the Teagasc Crops, Environment, and Land Use Programme.

Dr. Sana Arshad works at the Department of Geography at The Islamia University of Bahawalpur in Pakistan. Her research interests include environmental geography, climate change, and sustainable development practices in agriculture.

Dr. Firas Alsilibe is a faculty member in the Department of Transport Infrastructure and Water Resources Engineering at Széchenyi István University. His work involves the study of water resource management and climate change.

Dr. Atilla Vad is affiliated with the University of Debrecen's Institutes for Agricultural Research and Educational Farm. He is an expert in agricultural engineering.

Dr. Nagy Adrian also works at the University of Debrecen, focusing on Applied Economics and Sustainability.

Dr. Tamás Ratonyi is a researcher at the University of Debrecen's Institute of Land Use, Engineering, and Precision Farming Technology. His expertise lies in precision agriculture and land use management.

Prof. Dr. Manouchehr Gorji is a researcher at the University of Tehran, Department of Soil Science and Engineering.

Dr. Main Al-Dalahme is associated with the Hourani Center for Applied Scientific Research at Al-Ahliyya Amman University in Jordan. His research involves applied scientific research in various fields, including environmental science and sustainable development.

Dr. Safwan Mohammed is part of the Institute of Land Use, Engineering, and Precision Farming Technology at the University of Debrecen.