Methane enhancement and organic loading management of chicken manure and rice straw: Co-digestion synergistics and AI modelling through deep machine learning

By

Nehala Rauf

(Registration No: 00000403093)

Department of Energy Systems Engineering

US-Pakistan Centre for Advanced Studies in Energy (USPCASE)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

Methane enhancement and organic loading management of chicken manure and rice straw: Co-digestion synergistics and AI modelling through deep machine learning

By

Nehala Rauf

(Registration No: 00000403093)

A thesis submitted to the National University of Sciences and Technology, Islamabad,

in partial fulfillment of the requirements for the degree of

Master of Science in

Energy Systems Engineering

Supervisor: Dr. Muhammad Hassan

US-Pakistan Centre for Advanced Studies in Energy (USPCASE)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

THESIS ACCEPTANCE CERTIFICATE

Certified that the final copy of MS Thesis written by Mr / Ms <u>Nehala Rauf</u> (Registration No. <u>00000403093</u>), of <u>ESE, USPCASE</u> (School/College/Institute) has been vetted by the undersigned, found complete in all respects as per NUST Statutes/ Regulations/ Masters Policy, is free of plagiarism, errors, and mistakes and is accepted as partial fulfilment for the award of Master's degree. It is further certified that necessary amendments as pointed out by GEC members and foreign/ local evaluators of the scholar have also been incorporated in the said thesis.

Signature: _____

Name of Supervisor __Dr. Muhamad Hassan__

Date: __2 0-12-2024__

Signature (HOD): _____

Date: __24 -12 - 2024__

Signature (Dean/ Principal) _____

Date: __24 - 12 - 24__

National University of Sciences & Technology
MASTER'S THESIS WORK

We hereby recommend that the dissertation be prepared under our supervision by (Student Name & Regn No.) Nehala Rauf; 00000403093

Titled: **Methane enhancement and organic loading management of chicken manure and rice straw: Co-digestion synergistic and AI modelling through deep machine learning** be accepted in partial fulfilment of the requirements for the award of **MS Energy Systems Engineering** degree with (A) grade.

Examination Committee Members

1. Name: Dr. Rabia Liaquat   Signature: _____

2. Name: Dr. Mustafa Anwar   Signature: _____

3. Name: Dr Abeera Ayaz Ansari   Signature: _____

Supervisor's name: Dr. Muhammad Hassan   Signature: _____

Date: 04-12-2024

Dr. Naseem Iqbal
Head of Department   Signature   Date
                                                      24-12-2024

COUNTERSIGNED

Date: 24-12-24   Dean/Principal

US-Pakistan Center for Advanced Studies in Energy (USPCASE)

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in this thesis, entitled "Methane enhancement and organic loading management of chicken manure and rice straw: Co-digestion synergistics and AI modelling through deep machine learning" was conducted by Mr./Ms. Nehala Rauf under the supervision of Dr. Muhammad Hassan. No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the US-Pakistan Centre for Advanced Studies in Energy (USPCASE) in partial fulfilment of the requirements for the degree of Master of Science in the Field of Energy Systems Engineering Department of USPCASE National University of Sciences and Technology, Islamabad.

Student Name: Nehala Rauf   Signature: _____

Examination Committee:

a) GEC Member 1: Dr. Mustafa Anwar   Signature: _____
(Assistant Professor, ESE, USPCAS-E)

b) GEC Member 2: Dr. Abeera Ayaz Ansari   Signature: _____
(Assistant Professor, ESE, USPCAS-E)
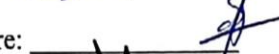
c) GEC Member 3: Dr. Rabia Liaqat   Signature: _____
(Associate Professor, ESE, USPCAS-E)

Supervisor Name: Dr. Muhammad Hassan   Signature: _____

Name of HOD: Dr. Naseem Iqbal   Signature: _____

Name of Principal/Dean: Dr. Adeel Waqas   Signature: _____

AUTHOR'S DECLARATION

I <u>Nehala Rauf</u> hereby state that my MS thesis titled <u>"Methane enhancement and organic loading management of chicken manure and rice straw: Co-digestion synergistics and AI modelling through deep machine learning"</u> is my work and has not been submitted previously by me for taking any degree from the National University of Sciences and Technology, Islamabad or anywhere else in the country/ world.

At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Name of Student:____Nehala Rauf_____

Date: _____20-12-2024_____

PLAGIARISM UNDERTAKING

I solemnly declare that the research work presented in the thesis titled "Methane enhancement and organic loading management of chicken manure and rice straw: Co-digestion synergistics and AI modelling through deep machine learning" is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and the National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above-titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above-titled thesis even after the award of MS degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and NUST, Islamabad have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student Signature: _____

Name: _____ Nehala Rauf _____

DEDICATION

This thesis is dedicated to my parents especially to my father; for their unwavering support, endless love and motivation throughout my academic journey; This work is a testament to their belief in me.

ACKNOWLEDGEMENTS

All praise and Gratitude are due to **ALMIGHTY ALLAH** the all-magnificent and all merciful, whose guidance and countless blessings have sustained me throughout my academic career.

I would like to express my heartfelt thanks to my dear supervisor **Dr. Muhammad Hassan**, whose expertise, guidance, and patience have been instrumental in shaping this thesis; Your insightful feedback has been invaluable and I'm deeply grateful for the learning experiences under your supervision.

An immense thanks to my GEC committee for their valuable guidance in designing the thesis.

A very heart-warming thanks to my colleagues and friends whose collaborations and fruitful discussions have supported me throughout this difficult time.

To the one who has extended his emotional support and guidance directly or indirectly; Thank you for making this journey possible.

TABLE OF CONTENTS

LIST OF TABLES

Page No.

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

CM   Chicken manure
RS   Rice straw
PM   Poultry manure
AVl   Applied Voltage
OT   Operating time
PT   Pretreatment Types
COD   Chemical oxygen demand
Pro   Proteins
Temp   Temperature
SMY   Specific methane yield

HRT   Hydraulic Retention Time
OLR   Organic loading rate
EI   Evaluation indicators
GBR   Gradient boosting regression

SVM   Support Vector Machine
RFR   Random forest regression
SVR   Support vector regression
XGBR   Extreme gradient boosting regression
PR   Polynomial Regression
CS   Corn straw
SB   Sugarcane bagasse
MG   Meadow grass
Car   Carbohydrates
VFA   Volatile fatty acids
FA   feed rate Amount
Oil C   Oil Content
TA   Total Alkalinity
TN   Total Nitrogen
TS   Total Solids
Cel   Cellulose
Hem   Hemicellulose
Lig   Lignin

SS   Soluble saccharide
CP   Crude protein
C/N   Carbon-Nitrogen ratio
S   Sulphur

ANN   Artificial neural network
DTR   Decision tree regression
KNNR   K-nearest neighbor regression
LR   Linear regression
CBR   CatBoost regression
PD   Poultry droppings
BS   briquette wheat straw
R^2 Determination Coefficient

ABSTRACT

The anaerobic digestion of chicken manure and rice straw presents a transformative opportunity for producing carbon-neutral biogas that aligns with global green energy initiatives. AD was extensively employed to remediate organic waste offering the dual benefits of generating renewable energy and nutrient-dense digestate side by side the process also helped to utilize the organic and agricultural waste in a very efficient way that followed the realm of waste to energy however, the process encounters instability which negatively impacts biogas production, to control and stabilize AD process machine learning has gained considerable attention in optimizing the process. Machine learning was used to predict the CH_4 (ml) with the application of Linear, Non-Linear and Ensemble learning regression models using sets of features including Days of digestion, pH, COD, TAN, FAN, TVFA as independent variables and CH_4 (ml) as a dependent variable obtained from the experimentation results of 70 days of digestion period in continuous stirring tank reactors for the process optimization of carbon-nitrogen ratio at different organic loading rate stages. The results demonstrated that C/N 24:1 was optimal for the efficient production of CH_4 under a continuous feeding rate. The study aims to propose the appropriate ML model from the comparison of 4 applied models from 70 samples in each reactor dataset; Important attributes are indicated by the pairwise Pearson correlation metrics heatmap. The ensemble learning models outperformed linear and non-linear regression models with a coefficient of determination (R^2) on training, testing and validation respectively 0.99, 0.96 and 0.84. Experimental results confirmed operational attributes revealed by Reactor D with the highest specific methane yield of 126.50% of the predicted value. Random Forest

feature importance elucidates total volatile fatty acid in Reactor A-B while pH in Reactor C-E is the important feature influencing the process. Ultimately, this research illustrates the efficacy of ML models for optimizing biogas production in AD providing valuable insights into improving the whole mechanism and enhancing methane yield from organic matter.

CHAPTER 1:   INTRODUCTION

1.1 Global Advancements and Environmental Challenges

The rapid pace of the worldwide population and innovative advancements increase the demand for commercialization and industrialization in developing countries [1]. Today's era of technological improvements, innovations and industrialization is advancing at an unprecedented rate. However, this progression reduces the commodity price thus increasing the demand for skillful employment, and increasing incomes. There is a tremendous surge of direct and indirect green jobs because of global renewable energy accelerating initiatives [2]. This speeding in population and developments that produce positive impacts on the world also cast escalating pollution, and rising fuel costs driven by continuous production and substantial energy consumption are demanding consideration issues [3]. Because of these advancements underdeveloped countries are facing severe water pollution, waste management and energy production crises. According to the current situation, out of many greatest hindrances in the establishment of the green economy, two major issues are wastewater treatment and cleaner energy production at meagre prices [4].

The main driving force behind globalization and industrialization is the energy sector if industry represents working towards national development, energy serves as the basic building block for sustaining the economy and implication of socio-economic advancements. Addressing these challenges requires innovative strategies to transition towards renewable energy systems. Due to huge socio-economic developments, the energy and food requirements also showed inclining trends during the last decade. Currently, the high energy prices and their fluctuations in international markets lead the researcher to explore further renewable and sustainable resources [5]. Electrical energy pricing fluctuations pose a substantial challenge in disrupting economic stability globally as well as in Pakistan. These pricing fluctuations are driven by vertical fiscal imbalance that complicates the sustainability of the energy sector [6].

1.2 Bioenergy Potential in Pakistan and Anaerobic Digestion

Addressing Pakistan's energy pricing through the incorporation of cleaner and more economical bioenergy resources can stabilize the energy pricing fluctuations. Pakistan has the potential to generate around 20,000 MW of electricity from biomass and 12615 million m3 of biogas annually [7]. Owing to massive sociodemographic developments, poultry meat consumption also gained favouritism resulting in millions of tons of poultry waste. and becoming a serious challenge for waste management authorities across the region [8,9]. Chicken manure has a high biogas production potential lab biomethane potential (BMP) has indicated that chicken manure has a biogas potential of methane 508ml CH4g-1VS when treated anaerobically [10]. Poultry manure that is rich in organic matter and microbes annual production of chicken manure discharge rate in Pakistan is 40000 million tons per year [11]. The chicken manure has a biogas power potential of 116GWh w.r.t its energy density at the rate of 2.14KWh/m3 [12].

Since Pakistan is an agricultural country so organic and anthropogenic waste production, segregation and management are other big problems that the country must mitigate immediately because these problems are causing severe health issues [13]. For the treatment of wastewater by anaerobic treatment process activated sludge made of animal manure and agricultural waste [14]. Rice straw (RS) is an organic substrate that can serve as a carbon source for microbial activity in AD this abundance makes it a viable feedstock for the AD process; Rice husk global production is 135.6 million tons [15]. In Pakistan, the average rice residue (rice husk + rice straw) production is almost 10.15 million tons which can be efficiently used to generate 44.227×106 MWh [16].

1.3 Anaerobic Digestion: A Synergistic Approach

Alternative advanced techniques should be implemented at a commercial scale that aims to promote sustainable development without degrading the environment. One possible synergistic solution to address both problems is anaerobic digestion (AD) a clean technology; is a multistage biological process that breaks down organic waste to produce nutrient-rich digestate and biogas. Converting biomass into valuable energy products enhances energy security, and this approach contributes to addressing the energy issue [17–19]. The detailed complex metabolic process driven by the synergistic microbes [20] is initiated in the anaerobic conditions by the hydrolytic microbes

when complex organic matter such as proteins, carbohydrates and lipids are converted into their simpler building units; amino acids, simple sugars and fatty acids respectively. Followed by the activity of acidogenic bacteria that convert soluble products into volatile fatty acids (VFA), hydrogen gas (H2) and Carbon monoxide (CO). Those intermediate products produced by acidogenesis are further broken down in acetogenesis by acetogenic bacteria; VFA into acetic acid and CO into CO2. But sometimes VFA accumulation can cause microbial activity inhibition. The final stage in the AD process is methanogenesis where methanogens (methane-producing bacteria) convert the prior phase intermediate products into methane and carbon dioxide [21].

Because of the complexity of the process it needs to be monitored deeply ML techniques have emerged as powerful tools in revolutionizing the way complex tasks; and they can be leveraged to analyze and model these complex interactions, leading to insights that can be used to optimize and to reduce the instability in the AD process [22]. The application of ML in the realm of AD holds significant potential to enhance the efficiency and performance of this biological process for organic waste treatment and biogas production [23].

1.4 Problem Statement

AD is a complex and time-expensive mechanism for biogas production its experimental monitoring of various parameters is not energy efficient and economically friendly, Advanced AI-based ML modelling would predict the optimal conditions and important features for enhanced methane production.

1.5 Research Objectives

This thesis aims to design the experiments of anaerobic digestion varied based on the C/N ratio in a continuous stirring tank reactor to collect the primary dataset later it will be preprocessed for detailed analysis and the main objectives of the thesis are:

- To perform the experimentation for identification of the optimal C/N ratio and OLR balance for enhanced methane production.

- To optimize the CSTRs C/N ratio and OLR balance, the primary dataset was collected from the real-time monitoring of CSTRs to evaluate the system efficiency for biogas production.

- To apply the predictive modelling using machine learning techniques to check performance stability.

- To identify the important features of AD through ML algorithms for the target variable methane gas volume based on the characteristics of the feedstock and seven features performance and developed ML models from real-time experimentation to generate predicted values for the optimized and efficient AD process.

CHAPTER 2:   LITERATURE REVIEW

2.1 Anaerobic Digestion and Machine Learning

AD reduces greenhouse gas emissions and environmental degradation by employing methane, a potent greenhouse gas, as a renewable energy source. likewise, it reduces the possibility of organic waste contaminating waterways [24]. The efficiency of this process is influenced by various factors, including the composition of the feedstock, operating conditions, and the dynamics of microbial communities [25,26]. The activated sludge of CM and RS [27] was co-digested in CSTR and can serve as the best alternative to enhance sustainability for renewable energy sources. Major significant AD driving factors impacting biogas production include organic loading rate (OLR), Carbon — Nitrogen ratio C/N and total volatile fatty acids (TVFA) assessed in this research work [28]. The C/N optimization of the sludge is an important attribute of the whole reaction mechanism. It is feasible to improve biogas yield and contribute to enhanced methane generation by optimizing the C/N [29]. One of the primary areas where ML can contribute to AD is in the prediction and control of process parameters. ML-trained models can analyze data from AD systems to predict the optimal conditions for maximizing biogas production. ML algorithms can analyze microbial data and key parameters to identify key microbial activity efficiency and their functions in the AD process [30].

2.2 Machine Learning Predictive Modelling

To optimize these features and minimize the experimental analysis error the real-time data processed through ML techniques that will enable a deeper observation of the process. ML is programming computers that involve training algorithms to learn from data interpret it and make predictions accordingly [31]. The integration of real-time data collection systems in anaerobic digesters provides a wealth of information that can be harnessed by ML algorithms. These algorithms can process large volumes of data, identify patterns, and make predictions, providing operators with actionable insights to make informed decisions [32]. This information can then be used to tailor the operating conditions to favour the growth and activity of beneficial microorganisms, leading to enhanced process stability and efficiency. In terms of feedstock

management, ML can assist in predicting the biodegradability of different organic materials [33]. By continuously monitoring a variety of parameters, ML models can identify patterns indicative of deviations from normal operation. This early detection allows for proactive intervention to address issues before they escalate, minimizing downtime and optimizing the overall efficiency of the AD system. The energy crisis and growing concern for effective waste management necessitates innovative solutions to efficiently convert organic waste into added-value products. The potential approach to deal with both problems is AD of CM and RS. To analyze this process the data collected from the experimentation was analyzed and visualized through a trained ML algorithm.

2.3 Working of Machine Learning Algorithms

2.3.1   Ensemble Learning Algorithms

The ensemble learning method works by combining the output of multiple trees each tree independently trains on a random subset of data and attributes and merges the output of each regression tree [34,35] by taking the average thus leveraging the robustness, enhancing prediction accuracy, and handling large amounts of non-linear datasets with high dimensions. In contrast to the decision tree, a random forest built with multiple trees during the training of data, the prediction of each tree is aggregated by taking the mean output of individuals which improves the performance and accuracy of a regression model [36]. Each tree in the forest is built from a random sample of data called bootstrapping. This randomness ensures the training of individual diverse trees on different subsets of data which reduce the correlation between trees and are less likely to overfit as decision trees.

2.3.2 Gradient Boosting Regression

Gradient Boosting Regression (GBR) is a robust ensemble ML technique utilized for regression purposes; it builds models by sequentially incorporating weak learners usually to rectify the errors of the prior models. Each succeeding tree is adapted to the residuals of the combined earlier trees, focusing on sections where the model is performing inadequately. Key elements of GBR comprise weak learners, additive modelling, a learning rate controlling the contribution of

each tree, a loss function for evaluating prediction precision, and regularization methods for combating overfitting. Although GBR exhibits elevated predictive accuracy and versatility in handling various data formats, fine-tuning hyperparameters such as the number of trees, tree depth, and the learning rate is necessary to prevent overfitting [37].

2.3.3 Extreme Gradient Boosting Regression

An advanced ML algorithm used for solving the regression problem renowned for its performance and speed is extreme gradient boosting regression. To increase overall prediction accuracy and speed performance, ensemble learning uses multiple base learners. This algorithm starts with the initial prediction and recursively adds trees for the residual prediction. The final prediction made by the algorithm based on the combined prediction of each subsequent tree focuses on the error made by its previous tree [38]. This methodology utilizes decision trees and gradient descent optimization in a combined form to construct a robust and effective predictive model.

2.3.4 Polynomial Regression

Polynomial regression is a statistical linear regression technique that employs a polynomial function to model the relationship between the target and feature variable. The complexity of the model is determined by the degree of the polynomial employed [39], with higher degrees presenting the risk of overfitting and lower degrees leading to underfitting, here in this research project PR with degree 2 is deployed.

2.3.5 Ridge Regression

Ridge regression is also a linear regression method that enhances ordinary least squares regression by incorporating a regularization term to address the issue of overfitting. This additional term, which penalizes large coefficient values, is calculated by summing the squares of the coefficients and multiplying them by a regularization parameter. The key objective of the model is to minimize the squared errors to reduce the coefficients towards zero. ridge regression reduces the risk of overfitting where variables show multicollinearity.

2.3.6 CatBoost Regression

The CatBoost regression is an advanced linear ML algorithm utilized for regression and classification objectives, specifically engineered to effectively manage and enhance model precision and efficiency known as CatBoost or categorical boosting. It belongs to the gradient boosting class and builds a group of decision trees sequentially, with each subsequent tree focusing on rectifying the residuals made by its predecessor trees. CatBoost is differentiated by its inherent capability to address categorical features directly, eliminating the requirement for extensive preprocessing, and employing a novel approach known as "ordered boosting" to prevent overfitting and ensure superior generalization. This methodology forms trees in a manner that minimizes prediction bias and variance.

2.3.7 K-Nearest Neighbor Regression

The K-Nearest Neighbors (KNN) regression method functions by estimating the target value for a new data point based on the 'k' nearest data points in the feature space from the training set. The algorithm initiates its work by representing each data point as a vector in multi-dimensional space and computes the distance between the existing and new points using the Euclidean distance metric. It identifies the 'k' closest neighbours, the data points with the shortest distances. The prediction for the new data point is then determined by averaging the target values of these 'k' nearest neighbours [40]. This approach is built on the assumption that data points that are similar will have comparable target values, thereby capturing local patterns in the data for accurate predictions.

CHAPTER 3:   METHODOLOGY

3.1 Material Collection and Methods

3.1.1 Material Collection and feedstock's chemical composition

The material Chicken Manure (CM) and Rice straw (RS) as biomass later that decomposed sludge was used in the CSTR. Fresh chicken manure was collected from Poultry control sheds in Sargodha, Punjab Pakistan and it was stored in a refrigerator at $2^{o}C$ for feeding at 15-day intervals. Wheat straw was collected from Punjab agricultural forms. It was first dried in an oven >$100^{o}C$ for 24 h, and then shredded, ground, and sieved to attain a uniform size of less than 1mm. The chemical composition of CM and RS was determined and reported in Table 1 (Chemical characterization of CM & RS): The inoculum was collected from a working biogas plant. After feeding CM, RS and inoculum in the CSTR daily dose of dextrose was fed at 2g/d for the sludge to activate and mature; this helped the methanogenic consortia to produce biogas efficiently. Lastly before feeding the sludge into CSTR to start the methane production process, it was strained by using 1mm sieves to remove all the unwanted substances from the sludge.

Table 3.1 designed to explain the chemical composition of CM, RS and seed sludge highlights distinct chemical compositions and characteristics for potential roles in AD. CM has a substantial amount of organic matter containing total solids (TS) and volatile solids (VS) 32.46% and 68.25% respectively. High nitrogen-enriched content reflects its potential for optimal AD activity; total nitrogen is 3.96% while total ammonia nitrogen (TAN) is in the concentration of 1080 mg/L. High nitrogen content led to an imbalance in the AD so C/N was carefully observed for the whole reaction mechanism CM itself has a low C/N ratio of ~9 and free ammonia nitrogen (FAN) was 49.68 mg/L further reinforcing the need for careful monitoring of microbial community for methane yield.

Table 3.1: Chemical composition of the chicken manure, rice straw and seed sludge

Category	Unit	CM	RS	Sludge
Total solids	%	32.46	93.12	2.16
Volatile solids	%	68.25	78.42	79.65
Ash	%	0.12	2.89	-
TN	%	3.96	0.96	1.14
TAN	mg/L	1080	-	1247
FAN	mg/L	49.68	-	53.68
TOC	%	36.78	49.68	12.57
Cellulose	%	-	41.57	-
Hemi-cellulose	%	-	19.68	-
Lignin	%	-	17.57	-
C/N	-	9.29	51.75	11.03
CODs	mg/L	4568	-	1057

In contrast, rice straw (RS) has a significantly higher TS and VS content of 93.12% and 78.42% respectively. RS is well known for its enriched carbon contents with total organic carbon (TOC) OF 49.68% and a high C/N ratio as well with a value of 51.75, which suggests it is well suited with CM to produce high methane yield. The lignocellulosic composition of RS comprised cellulose, hemicellulose and lignin with 41.57%, 19.68% and 17.57% respectively. Seed sludge provides a steady microbial inoculum with suitable nutrients including TS, VS, TN, C/N, TOC content, TAN and FAN of 2.16%, 79.65%, 1.14%, 11.03, 12.57%, 1247mg/L and 53.68mg/L respectively to optimize the system.

3.2 Experimental C/N Optimization

To address the challenge sludge was derived from chicken manure (CM) which is a nutrient-dense and biologically active sludge that served as the cornerstone of this approach; the methodology is designed to maximize the efficiency of AD. The activated sludge of CM and RS was co-digested in CSTR for 70 days in five phases. Each phase has 15 days of hydraulic retention time, five different organic loading rates, and effluent removal. The CSTRs used in this research have the capacity of 10L working volume, they are adapted from the previous research [29]. The Anaerobic Co-digestion process was carried out in 5 CSTRs named A-E based on different C/N ratios, C/N was optimized for all 5 reactors. Reactor A operated on the sludge with C/N 15:1, reactor B was operated with C/N of 18:1, reactor C operated with C/N of 21:1, reactor D was operated with C/N of 24:1 and reactor E operated with C/N of 27:1. Each CSTR was operated in

(I-V) phases based on 5 different organic loading rate (OLR) or feeding concentration. The hydraulic retention time (HRT) for the daily feeding and slurry removal was 15 days for phases (I-VI) and 10 days for phase V; HRT for all reactors were kept the same for each phase to make the comparison of methane gas production in different C/N ratios. In phase I for the 1st 15 days of digestion CSTR 1.5VS/L. d was loaded. In phase II from the 16th - 30th day, 3.0 VS/L·d was fed into CSTR. In phase III 31st - 45th days of digestion the daily feeding rate was 4.5 VS/L·d. In phase-IV 6.0 VS/L·d was loaded from the 46th – 60th day, Finally, phase-V has a feeding of 7.5 VS/L·d from the 61st- 70th days of the digestion period. For pseudo-state steady stabilization; CSTRs were operated one week after each phase of applied OLR. Biochemical operational parameters like COD, TAN, FAN, TVFA, pH and methane production were determined daily.

3.3 Analytical Technique and Procedures

The chemical parameters of the sludge such as pH, COD, TVFA, TAN, FAN and biogas contents including methane production were determined. The pH was measured with a HI9829 multiparameter pH waterproof meter with a GPS option from HANNA instruments with a pH accuracy of ±0.02. Total volatile fatty acids and alcohol concentrations in the CSTR's digestate were determined by (GC-2014, Shimadzu, Japan) with thermal conductivity detector specification (DA-Stabilwax, 30 m × 0.53 mm × 1 l m) Biogas contents were measured by Gas chromatography (GC-2010 Pro, SHIMADZU Japan) equipped with a TCD column (RT-MS5A, 30 m × 0.32 mm ID, 30 μm) for the detection of CH_4. The COD removal efficiency was determined; In each phase of the AD experimentation, the change in COD removal efficiency was measured, CODi is the initial chemical oxygen demand and CODf is the final chemical oxygen demand [29].

$$CODs\ change\% = \frac{COD_i - COD_f}{COD_i} \times 10 \qquad (3.1)$$

$$FAN = TAN \left[1 + \frac{10^{-pH}}{10^{-\left(0.09018 + \frac{2729.92}{T(k)}\right)}}\right]^{-1} \qquad (3.2)$$

Where FAN is the concentration of free ammonia nitrogen in mg/L, TAN is the concentration of total ammonia nitrogen was measured by (Lian-hua Tech Co., Ltd. China) and Temperature (T) measured in kelvin (k) [41].

3.4 Data Collection; Interpretation and ML Understanding

The approach intended to develop a model to predict the behaviour of AD with the help of ML techniques; to obtain the purpose, a laboratory setup was organized for AD, where C/N optimization in 5 different reactors was examined. For the analysis, data were stored during the data gathering. The input dataset was prepared from the operating results of 5 laboratory scale CSTRs that were run for 70 days with a working volume of 10L were operated in a continuous mode at five different phases with a gradual increase in the OLR starting from 1.5 VS/L·d to 7.5 VS/L·d for every reactor at 5 different C/N ratios to predict the optimized methane production. For each reactor, the output data were collected with the help of a methane sensor and by the process of water displacement.

For data preparation, multiple spreadsheets that corresponded to daily production were produced and then merged into a single data frame to deploy regression-based models. The 2nd step after the data collection in the data cleaning includes removing the duplicate values and correcting errors (anomalies) if any. The 3rd step of the process is data understanding and visualization it includes summary statistics (mean, mode, medians, etc to understand the distribution of data) illustrated in figures 3.1-3.10 and whether the data is numerical or categorical [42].

The 4th step is data visualization and correlation analysis which is explained by Pearson pairwise correlation heatmaps which explains the pairwise correlation of each variable with another and generates a specific value against these two variables. After that, the whole data is partitioned randomly into 3 subsets named training, testing, and validation after this dataset is ready for regression-based model building, training, and deployment, the pipeline of the whole process is explained in Figure 3.11 For the implementation of ML, the raw restructured and converted into a

Figure 3.1 Boxplot distribution of features in Reactor A

Figure 3.2 Methane production values over time of Reactor-A

Figure 3.3 Boxplot distribution of features in Reactor B

Figure 3.4 Methane production values over time of Reactor-B

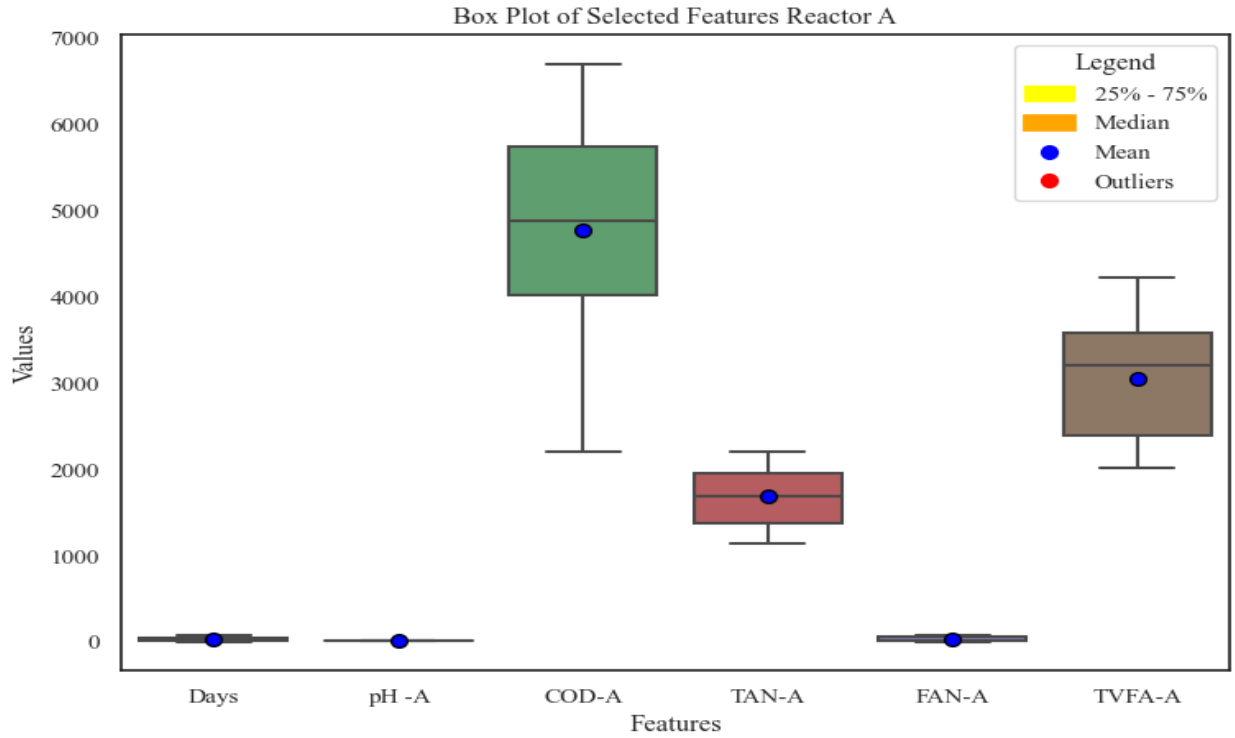**Figure 3.5** Boxplot distribution of features in Reactor C

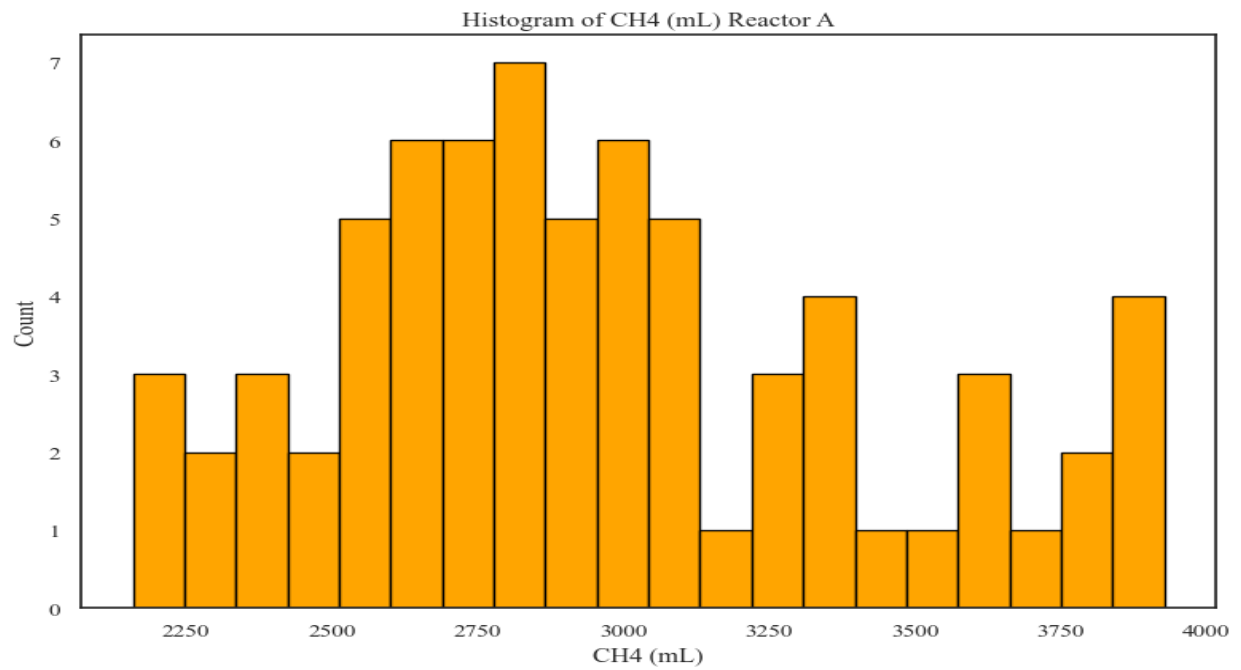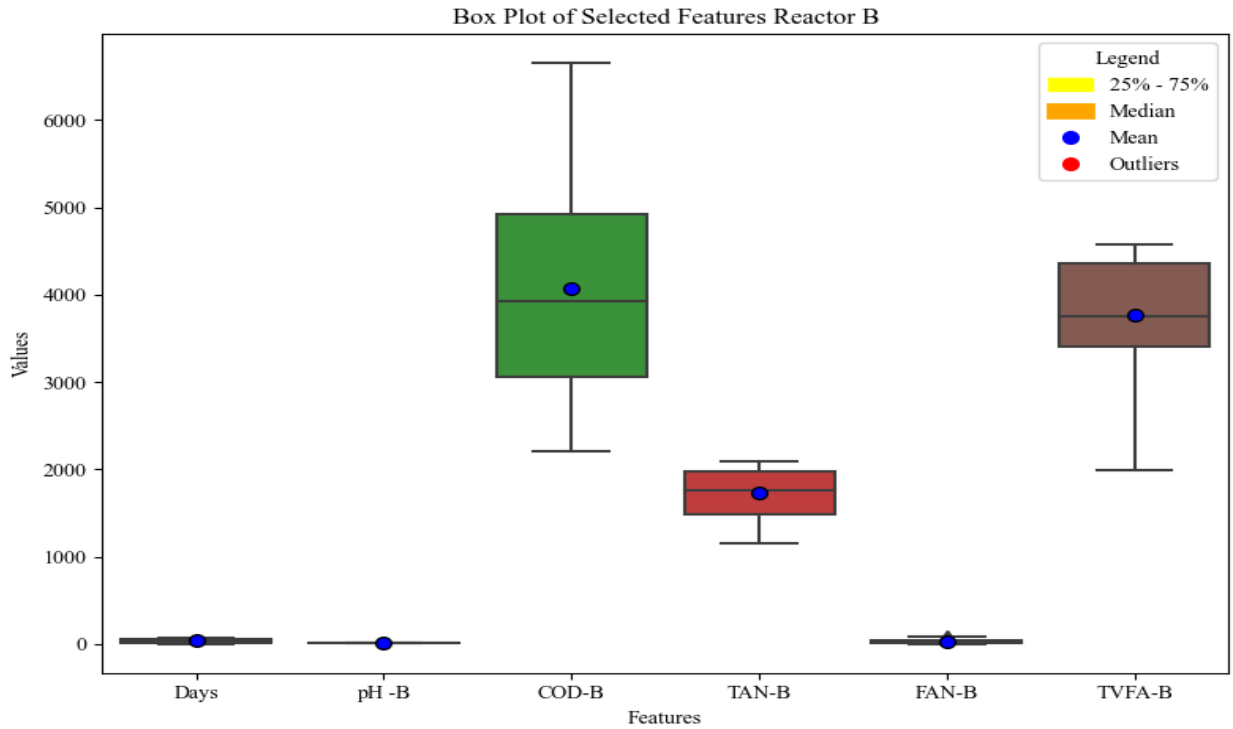**Figure 3.6** Methane production values over time of Reactor-C

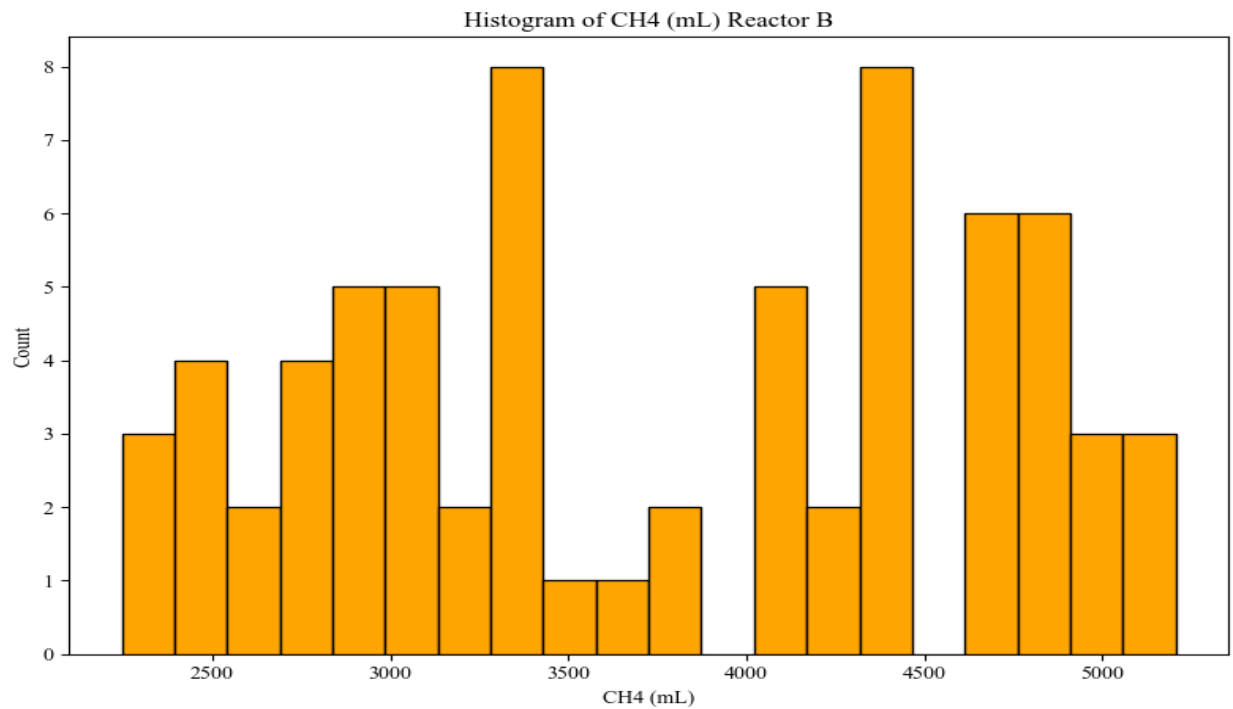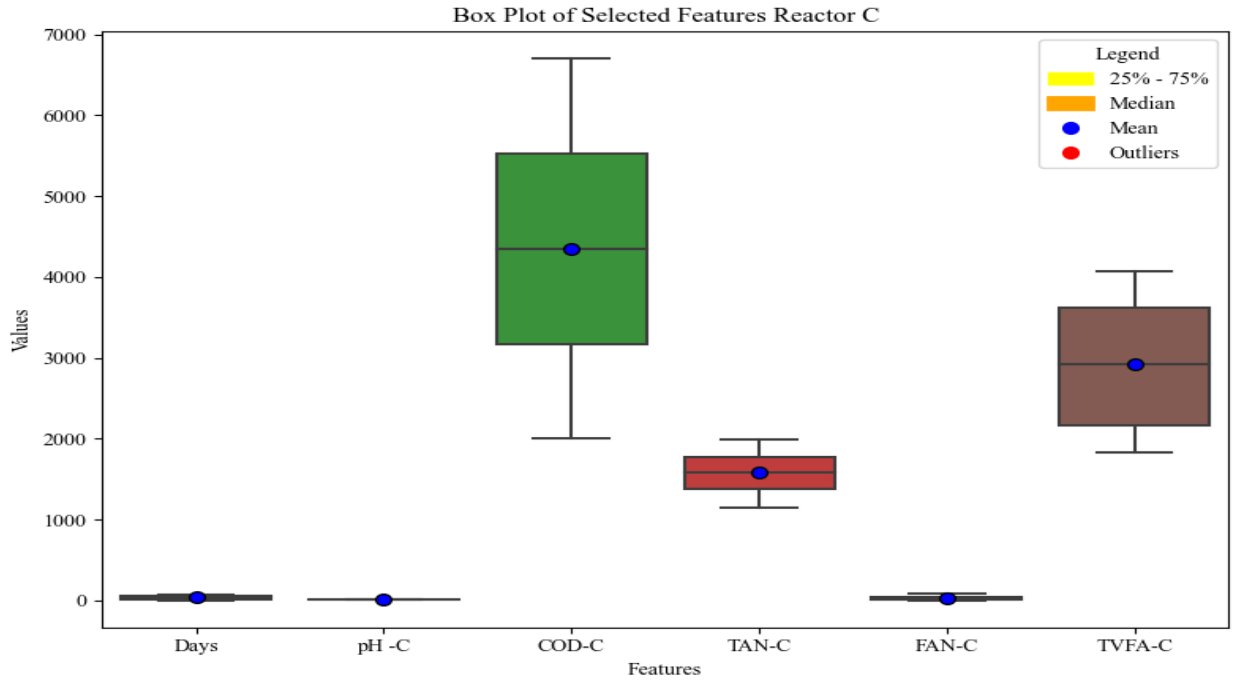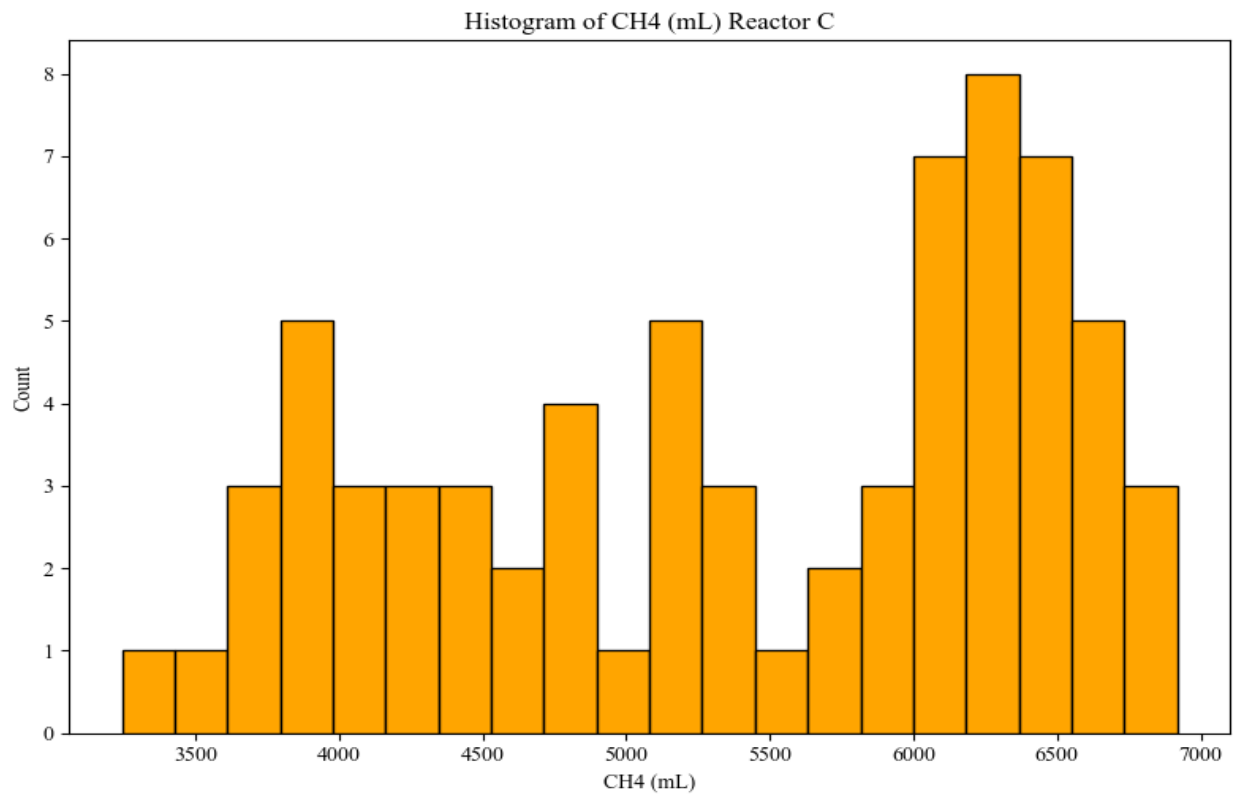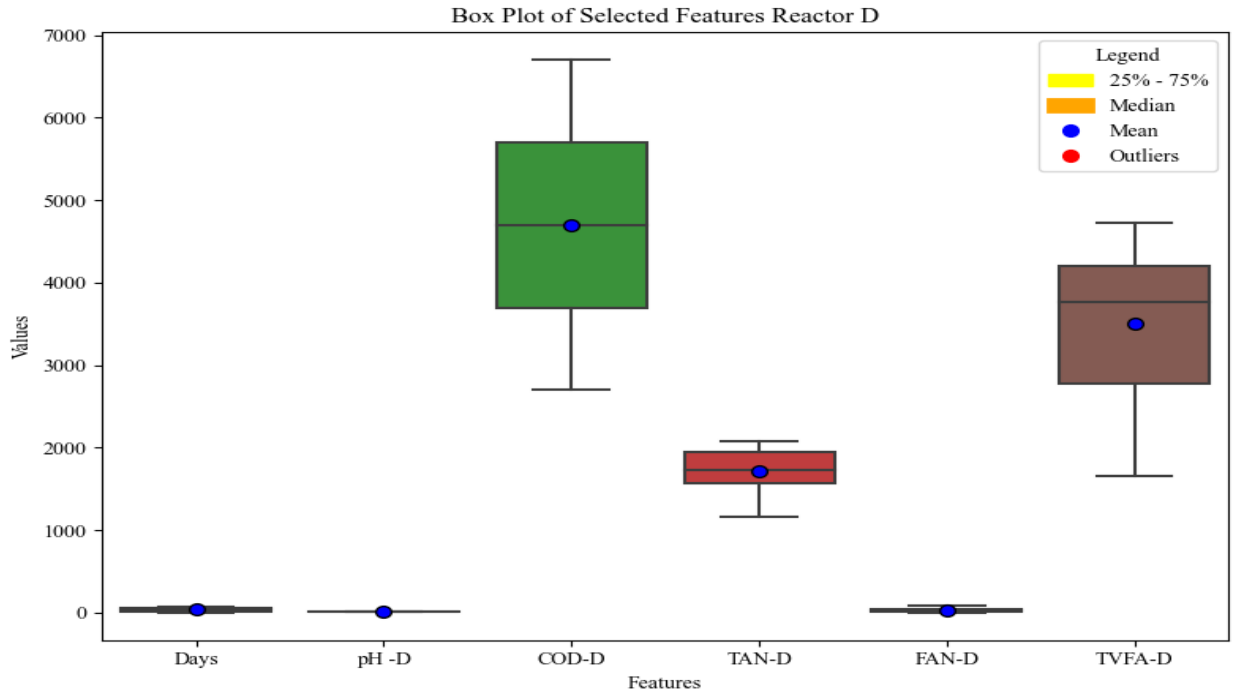**Figure 3.7** Boxplot distribution of features in Reactor D
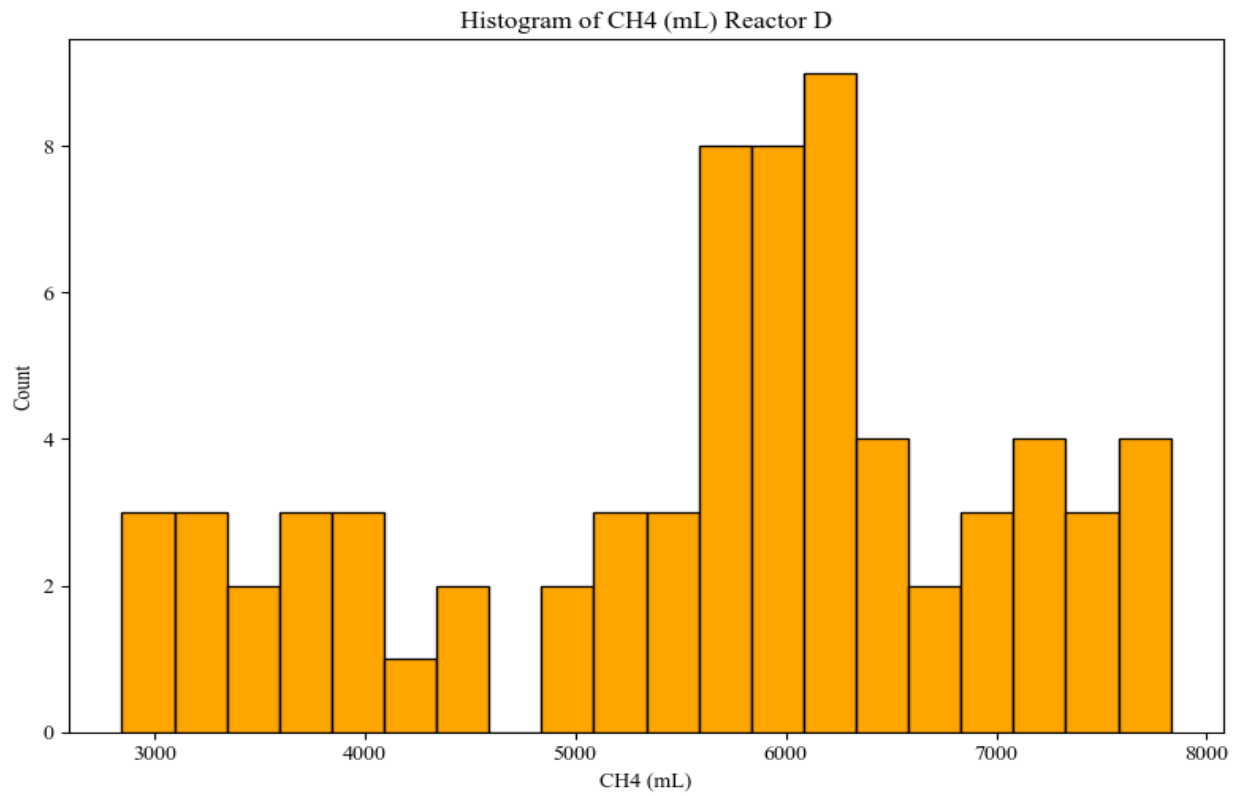
Figure 3.8 Methane production values over time of Reactor-D
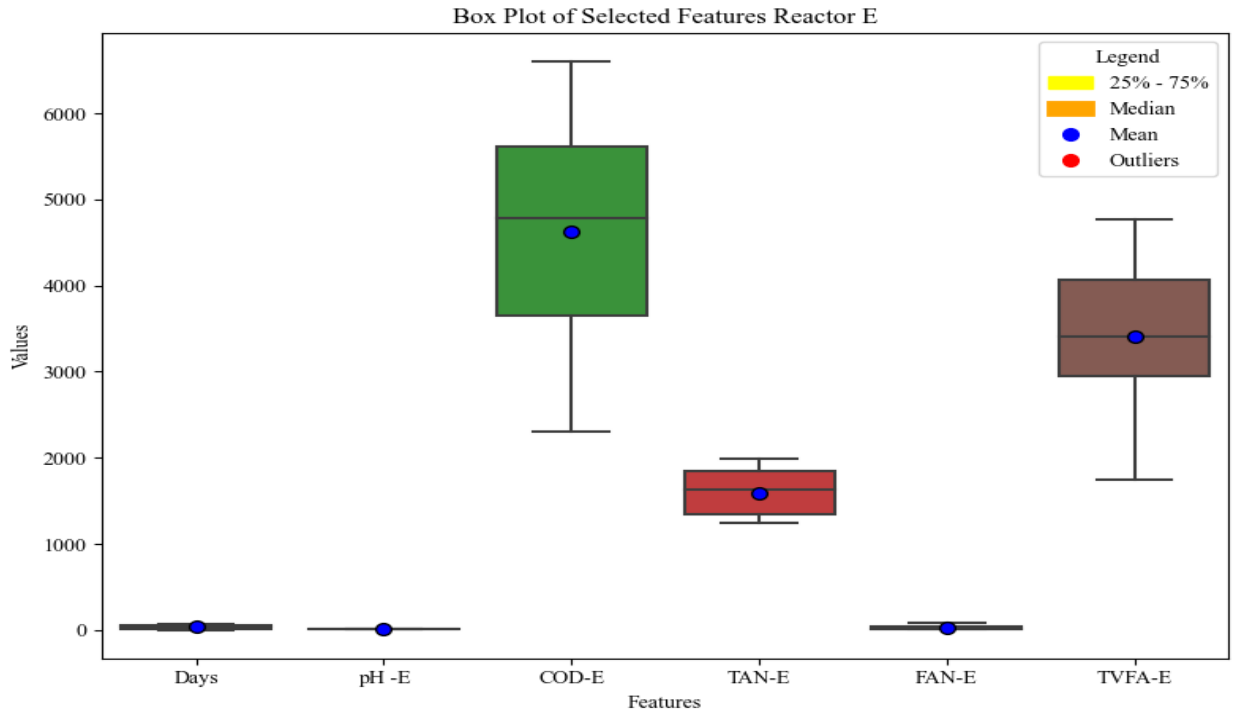
Figure 3.9 Boxplot distribution of features in Reactor E
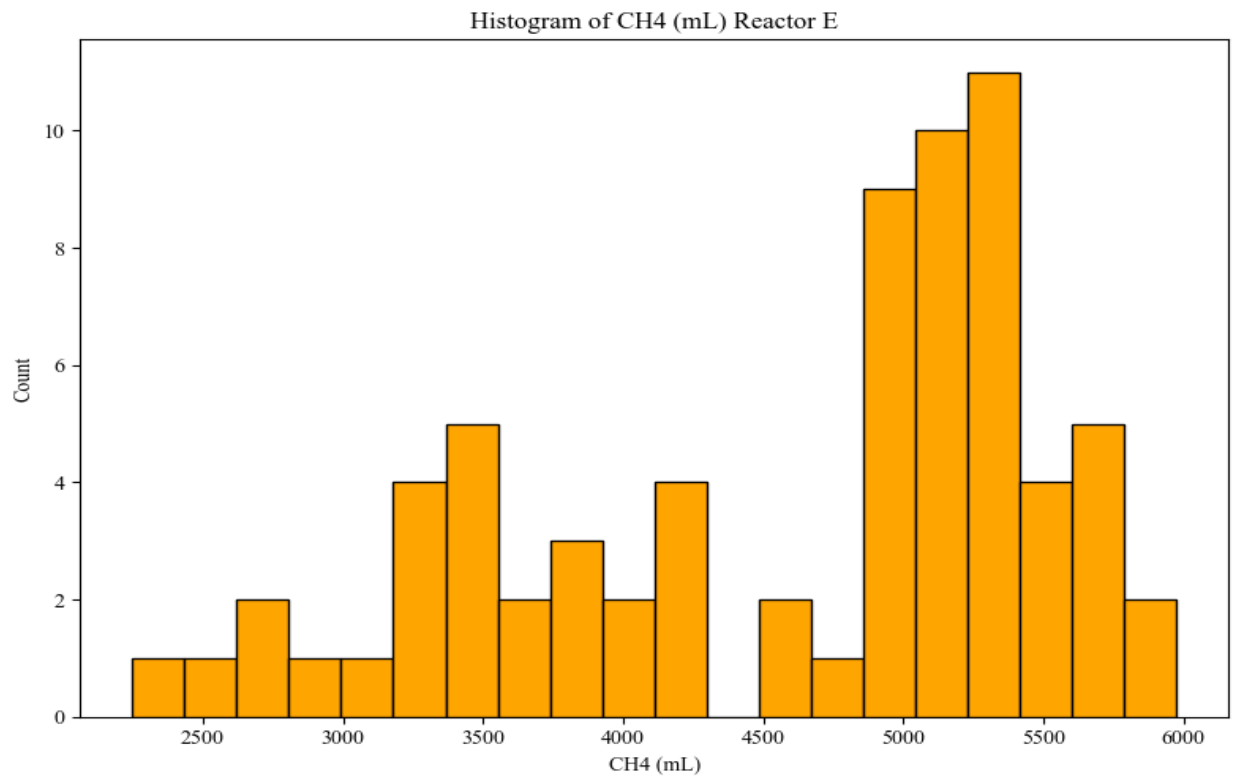
Figure 3.10 Methane production values over time of Reactor-E

Table 3.2: Statistical properties of a dataset of Reactor-A

	Days	pH -A	COD-A	TAN-A	FAN-A	TVFA-A	CH4 (mL)
count	70.000000	70.000000	70.000000	70.000000	70.000000	70.000000	70.000000
mean	35.500000	6.976429	4765.000000	1690.271429	30.446884	3055.971429	2963.975900
std	20.351085	0.398239	1243.427649	311.914143	26.007845	616.957424	463.881405
min	1.000000	6.400000	2200.000000	1142.000000	3.888760	2012.000000	2158.000000
25%	18.250000	6.600000	4012.500000	1381.750000	7.662788	2389.000000	2646.493750
50%	35.500000	6.940000	4875.000000	1688.000000	19.799330	3210.000000	2879.880000
75%	52.750000	7.335000	5737.500000	1949.250000	53.839100	3586.500000	3281.540000
max	70.000000	7.600000	6700.000000	2208.000000	81.157630	4226.000000	3927.680000

Table 3.3: Statistical properties of a dataset of Reactor-B

	Days	pH -B	COD-B	TAN-B	FAN-B	TVFA-B	CH4 (mL)
count	70.000000	70.000000	70.000000	70.000000	70.000000	70.000000	70.000000
mean	35.500000	6.976429	4075.000000	1723.828571	27.851681	3769.942857	3744.130764
std	20.351085	0.398239	1243.060447	269.106356	23.015676	628.085102	894.147314
min	1.000000	6.400000	2200.000000	1157.000000	3.763090	1986.000000	2245.000000
25%	18.250000	6.600000	3062.500000	1489.250000	10.476100	3403.250000	2919.341250
50%	35.500000	6.940000	3925.000000	1765.000000	18.796750	3752.000000	3728.240000
75%	52.750000	7.335000	4925.000000	1976.750000	40.402100	4365.500000	4612.660000
max	70.000000	7.600000	6650.000000	2095.000000	88.277570	4578.000000	5206.500000

Table 3.4: Statistical properties of a dataset of Reactor-C

	Days	pH -C	COD-C	TAN-C	FAN-C	TVFA-C	CH4 (mL)
count	70.000000	70.000000	70.000000	70.000000	70.000000	70.000000	70.000000
mean	35.500000	6.976429	4349.999986	1579.900000	28.084031	2919.928571	5391.777571
std	20.351085	0.398239	1386.233292	237.305658	23.702160	753.453697	1052.137574
min	1.000000	6.400000	2000.000000	1146.000000	3.544280	1824.000000	3245.000000
25%	18.250000	6.600000	3175.000000	1382.250000	10.169335	2165.750000	4456.250000
50%	35.500000	6.940000	4350.000000	1588.000000	17.812655	2925.000000	5627.240000
75%	52.750000	7.335000	5525.000000	1775.000000	44.371552	3618.250000	6356.500000
max	70.000000	7.600000	6700.000000	1995.000000	83.854150	4068.000000	6917.250000

Table 3.5: Statistical properties of a dataset of Reactor-D

	Days	pH -D	COD-D	TAN-D	FAN-D	TVFA-D	CH4 (mL)
count	70.000000	70.000000	70.000000	70.000000	70.000000	70.000000	70.000000
mean	35.500000	6.976429	4700.000000	1709.114286	27.900330	3504.600000	5652.752743
std	20.351085	0.398239	1179.773052	255.909718	20.560387	906.644768	1364.915263
min	1.000000	6.400000	2700.000000	1157.000000	3.939840	1658.000000	2845.000000
25%	18.250000	6.600000	3700.000000	1568.000000	10.530255	2773.000000	4996.250000
50%	35.500000	6.940000	4700.000000	1730.500000	24.340195	3763.000000	5922.840000
75%	52.750000	7.335000	5700.000000	1943.500000	40.523880	4200.000000	6537.605000
max	70.000000	7.600000	6700.000000	2078.000000	81.157630	4725.000000	7828.710000

Table 3.6: Statistical properties of a dataset of Reactor-E

|	Days	pH -E	COD-E	TAN-E	FAN-E	TVFA-E	CH4 (mL)
Count	70.000000	70.000000	70.000000	70.000000	70.000000	70.000000	70.000000
Mean	35.500000	6.976429	4619.757143	1594.514286	27.986286	3408.500000	4579.745214
Std	20.351085	0.398239	1241.989480	250.834996	24.373529	844.535766	954.836669
Min	1.000000	6.400000	2310.000000	1247.000000	3.990000	1745.000000	2248.000000
25%	18.250000	6.600000	3657.500000	1345.000000	7.600000	2954.000000	3800.000000
50%	35.500000	6.940000	4788.500000	1629.000000	17.455000	3411.500000	5007.655000
75%	52.750000	7.335000	5621.000000	1842.500000	43.087500	4067.250000	5334.725000
Max	70.000000	7.600000	6600.000000	1985.000000	81.210000	4768.000000	5974.360000
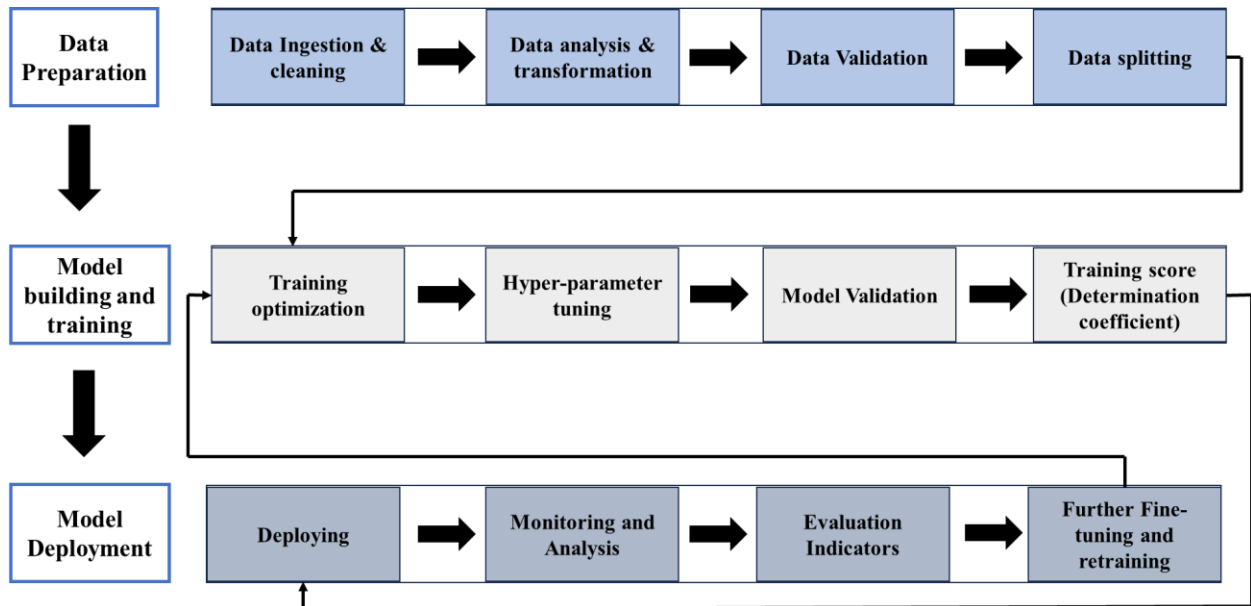
Figure 3.11 Understanding of the Machine Learning process from data collection to algorithm deployment

suitable format for the ML application, and the following ML models were deployed and evaluated, Python was the programming language used for this research and ANACONDA Navigator was the programming environment that provides various libraries essential for ML such as sci-kit, learn to import the ML algorithms, Pandas and Matplotlib for the visualization and plotting of graphs. Evaluation indicators for regression models are essential to understanding the performance and accuracy of prediction made by the model. The coefficient of determination denoted as R^2, measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It is a key indicator of the explanatory power of a regression model. An R^2 value of almost equal or less than 1 indicates that the model explains none of the variability of the response data. A higher R^2 value signifies a better fit, meaning the model effectively captures the relationship between the dependent and independent variables [43].

$$R^2 = \frac{\sum(y, -\hat{y}_1)^2}{\sum(y_1 - \bar{y})^2} \qquad (3.3)$$

CHAPTER 4:   RESULTS AND DISCUSSION

4.1 Methane production profiles in all CSTRs and their correspondent anaerobic digestion parameters: CSTRs stability

In the observed CSTR running an AD process over 70 days, methane production initially increased as the organic loading rate (OLR) was progressively elevated from 1.5 to 6.0 Volatile Solids per Liter per Day (VS/L/D) through four sequential 15-day phases. This upward trend in methane generation signifies that the microbial community within the reactor efficiently adapted to and metabolized the increased higher amounts of substrates, optimally converting them into methane. However, a notable shift occurred in the final phase when the OLR was increased to 7.5 VS/L/D but over a shorter span of 10 days. During this phase, methane production unexpectedly decreased, highlighting a critical point where the shortened adaptation period overwhelmed the microbial ecosystem. This overloading likely led to process imbalances, such as the accumulation of intermediates that methanogens couldn't efficiently convert to methane, thereby impacting overall methane yield negatively. Table 7 elaborated; that this research aimed to optimize AD by adjusting the C/N ratio; a critical parameter to assess process efficiency. The data indicates methane yield enhancement in varying C/N ratios, In C/N 15 the methane yield is relatively modest at 54.21%. While discussing C/N 18 the nutrient balance for maximizing microbial activity is relatively increasing at 99.73% the enhancement reflects a more favourable carbon and nitrogen balance which supports more efficient microbial metabolism. However, with a further increase in C/N to 21 there is a slight decrease in the yield of methane valued at 89.09%.

The most substantial increase in the yield of methane occurs at C/N 24 which is 126.50% this C/N concentration is ideal and promotes maximum microbial efficiency and methane output. While methane yield decreases to 83.36% at C/N 27 thus the further increase in C/N would deviate the reaction from optimal conditions. While comparing current studies with the literature on other feedstock compositions and methane yield there is a significant increase in the yield of methane with 250.3% when the mixing ratio of CM and RS is equal, far exceeding the yield with varying concentrations of C/N may provide a synergistic effect for optimizing AD [44]. When CM or

poultry manure was mixed with other agricultural residues like corn stover with feedstock composition of CM: corn stover 1:3 methane yield was 298.2% [45]. On the other hand, with the

Table 4.1: Comparison of previous and current studies with different C/N in CSTRs

Feedstock	Mixing ratio (VS based)	Methane Yield (%)	References
CM: RS	C/N~15	54.21	Current studies
CM: RS	C/N~18	99.73	Current studies
CM: RS	C/N~21	89.09	Current studies
CM: RS	C/N~24	126.50	Current studies
CM: RS	C/N~27	83.36	Current studies
CM: RS	1:1	250.3	[44]
PM: CS	1:3	177.6	[46]
PM: corn stover	1:3	298.2	[45]
PD: SB: press mud	3:1:1	56.9	[47]
PD: BS	1 :1	56.9	[48]
PD: MG	1 :1	57.6	[49]

the same mixing ratio of CM and corn straw methane yield was 177.6% [46]. Conversely, some feedstock compositions of poultry manure with briquette wheat straw and meadow grass yield very little methane at 56.9% and 57.6% respectively [48,49].

4.2 Effect of input parameters on the target methane gas

The observed trends in the CSTRs during AD demonstrated in Figure 4.1 highlight the dynamic nature of microbial community processes and their influence on methane production. The initial rise in pH followed by a decline suggests changing the reactor's AD phases of substrate digestibility. Monitoring pH remains critical for understanding and optimizing AD processes for sustainable biogas production [50]. pH is an actual important variable proved by experimentation of varied C/N ratios as well as ML. The fluctuation in pH can lead to the formation of dead zones of microbial community which will ultimately cause the reaction to cease. Methane production is influenced by the rate and extent of COD reduction as the degradation of the organic matter provides the substrate for methanogenesis. All the reactors have a decreasing COD trend reflecting the stability of the AD [51]. Reactors A, B, C, D, and E have the measured COD values in the initial stage of the reaction ranging from 6600-6700 mg/L while the COD values range in the 3rd month of the reaction 2000-2700 mg/L. Monitoring these COD trends throughout the AD period is therefore vital, providing insights into organic matter degradation rates, biogas production

potential, and the health of the microbial community. Monitoring total ammonia nitrogen (TAN) trends is essential for estimating AD efficiency and the complicated nutrient dynamics within the system [52]. The increasing trend in TAN observed over the 70 days of anaerobic CSTRs ranging from 1150-2200 mg/L indicates the progressive accumulation of ammonia and ammonium compounds. This accumulation is primarily from the breakdown of nitrogen-rich organic compounds, such as proteins into amino acids by the diverse microbial community within the reactors contributing to the rising TAN levels. The free ammonium nitrogen (FAN) trend is also increasing up to the 40th operational day of AD then followed by a gradual decrease suggesting stabilization of the microbial community. Managing nitrogen levels becomes essential to maintain process stability and optimize biogas production. The increasing trend in Total Volatile Fatty Acids (TVFA) observed in CSTR identified processes and organic matter breakdown during the initial stages of AD. As organic compounds are degraded, VFAs accumulate often as a result of the imbalance between acidogenic and acetogenic bacterial activity, when VFAs are produced faster than they are consumed and acetogenic bacteria are lacking [53], VFAs level rises indicating that the microbial population is adjusting the process conditions; reaching their peak value on the 45th-50th day in the digestion period. Subsequently, the slight decline of TVFA levels suggests the utilization of these acids by microbial populations for energy and the conversion into methane and carbon dioxide. This dynamic trend underscores the intricate balance between VFA production and consumption within the system. Monitoring TVFA trends provides valuable insights into microbial activity [54], process stability, and biogas production potential, aiding in the optimization of the process.
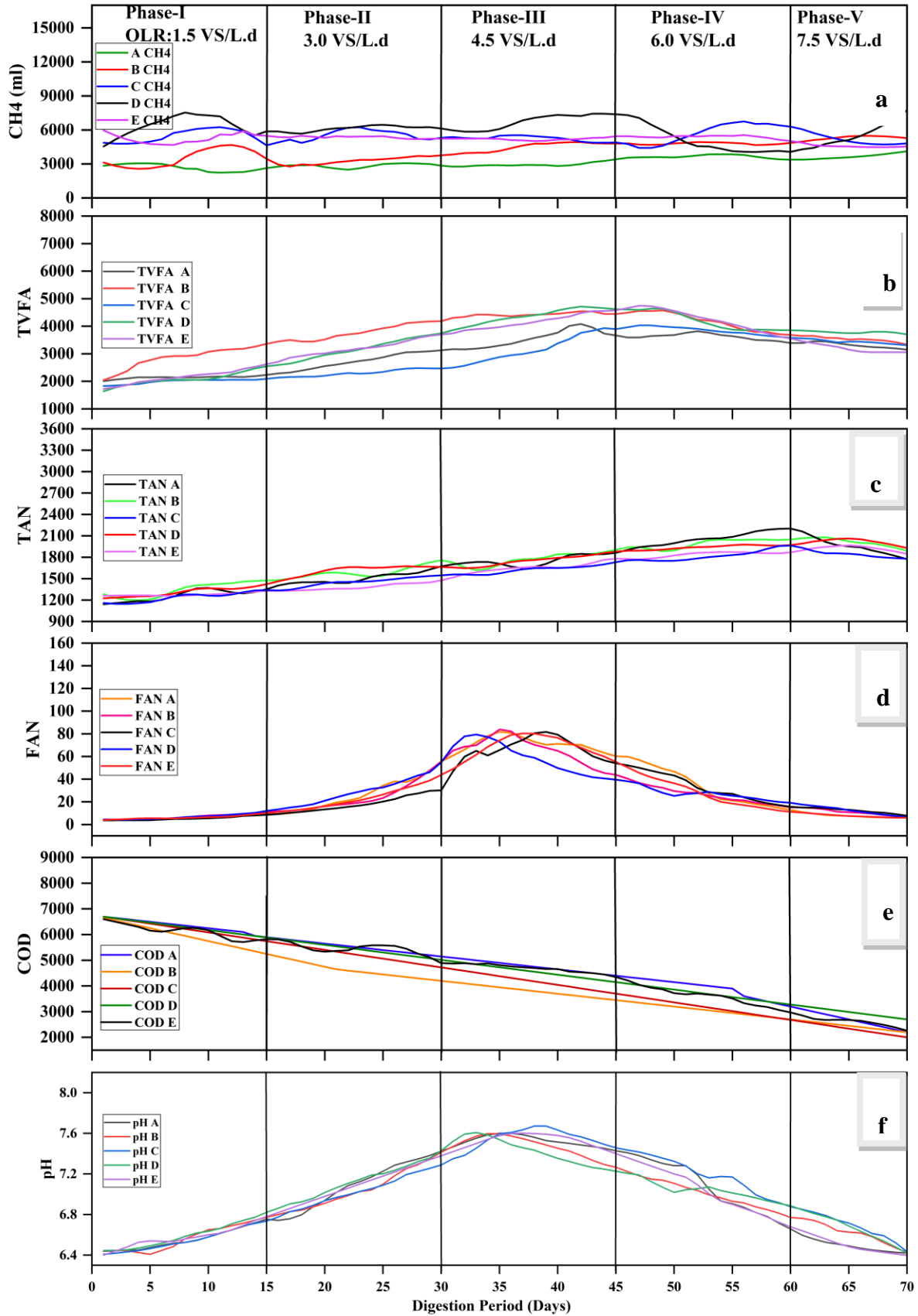
Phase-I
OLR:1.5 VS/L.d

Phase-II
3.0 VS/L.d

Phase-III
4.5 VS/L.d

Phase-IV
6.0 VS/L.d

Phase-V
7.5 VS/L.d

a

- A CH4
- B CH4
- C CH4
- D CH4
- E CH4

b

- TVFA A
- TVFA B
- TVFA C
- TVFA D
- TVFA E

c

- TAN A
- TAN B
- TAN C
- TAN D
- TAN E

d

- FAN A
- FAN B
- FAN C
- FAN D
- FAN E

e

- COD A
- COD B
- COD C
- COD D
- COD E

f

- pH A
- pH B
- pH C
- pH D
- pH E

Digestion Period (Days)

25

Figure 4.1 :(a) methane production dataset across 70 days of co-digestion in 5 OLR stages of Reactors A-E (b) TVFA trend of Reactors A-E in 5 different OLR stages (c) TAN trend of Reactors A-E in 5 different OLR stages (d) FAN trend of Reactors A-E in 5 different OLR stages (e) COD trends of reactors A-E in 5 different OLR stages (f) pH trends of Reactors A-E in 5 different OLR stages

4.3. Pearson correlations dynamics of different attributes in CSTRs: Understanding of co-relations and modelling of CSTRs by machine learning

In ML a pairwise Pearson correlation metric is an essential visualization technique that helps to analyze and interpret data by examining the linear correlations between each dependent variable itself and the independent variable [55]. In the datasets of the reactors, the correlation heatmap reveals the intricate correlation between the positive and negative values of all the feature variables with themselves and with the target variable methane gas revealed in Figure 4.2. Moreover, it aids in feature selection by pointing out redundant features with high correlation values that may indicate information overlap. Elimination of these redundant feature variables, the model is simplified, reducing the risk of overfitting can be a substantial issue for regression-based machine ML algorithms and enhancing overall efficiency [56,57]. This multi-stage process is directly proportional to methane production and the correlation metrics relevance investigates how it helps with the inability to separate highly correlated variables detection and overall model interpretation. The carbon-to-nitrogen ratio (C/N) is an important attribute of the AD process since it determines the balance of nutrition, in this research C/N were kept in the range of 15-27 for Reactor A, B, C, D and E with a respective difference of 3 in each reactor condition. Previous studies have shown that the C/N standard range for optimum biogas production is between 20-30 lower C/N may inhibit the process because of the plentiful free residual ammonia gas, If the substrate contains poultry manure, the methanogen will utilize nitrogen and retard the AD mechanism [58]. In the first stage, oxygen is gradually depleting until the last stage of AD, the methanogenic stage, where methanogen produces biogas that contains up to 70% methane gas, and at that particular time, the soluble chemical oxygen demand is inadequate, indicating maximum biogas production. At the start of the reaction in the hydrolysis phase, rapid methane production is not feasible within the reactors and the same case is explained in experimentation; a visible increase in methane production starts from acidogenesis and acetogenesis.
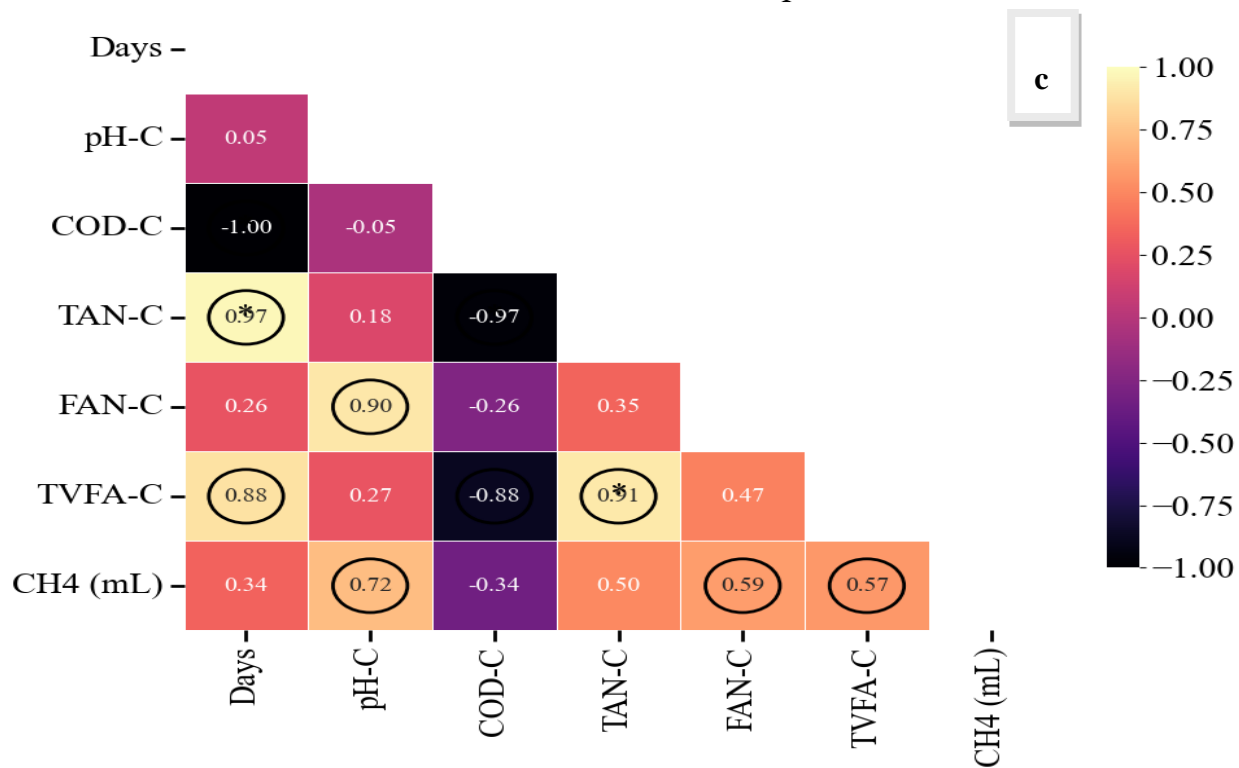
26

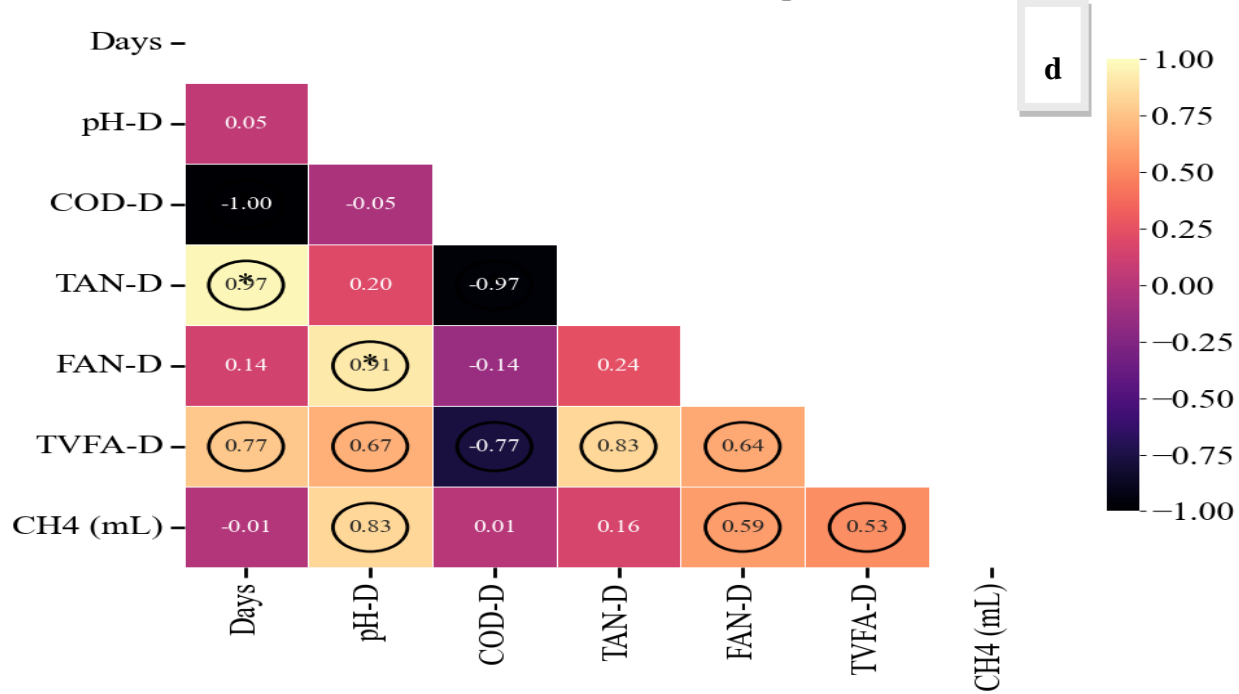Pearson Pairwise Correlation Heatmap of Reactor-A
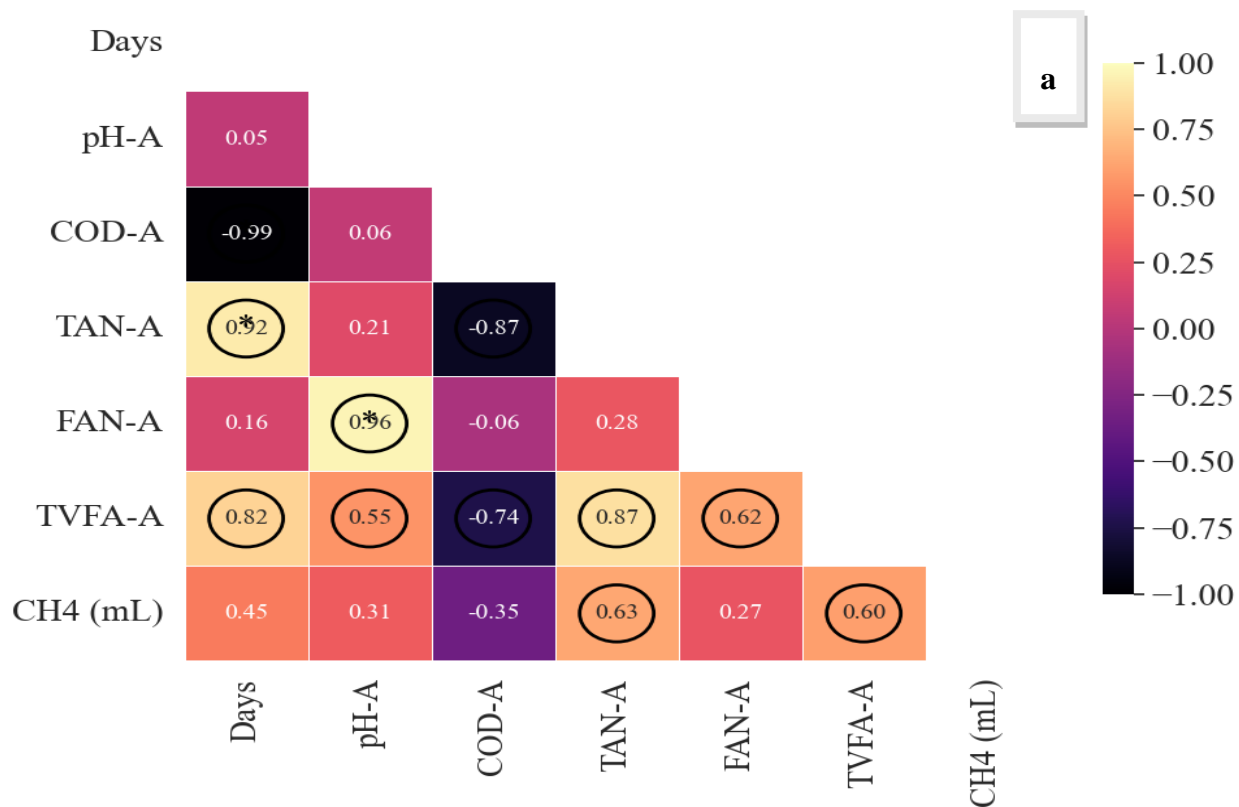
Pearson Pairwise Correlation Heatmap of Reactor-B
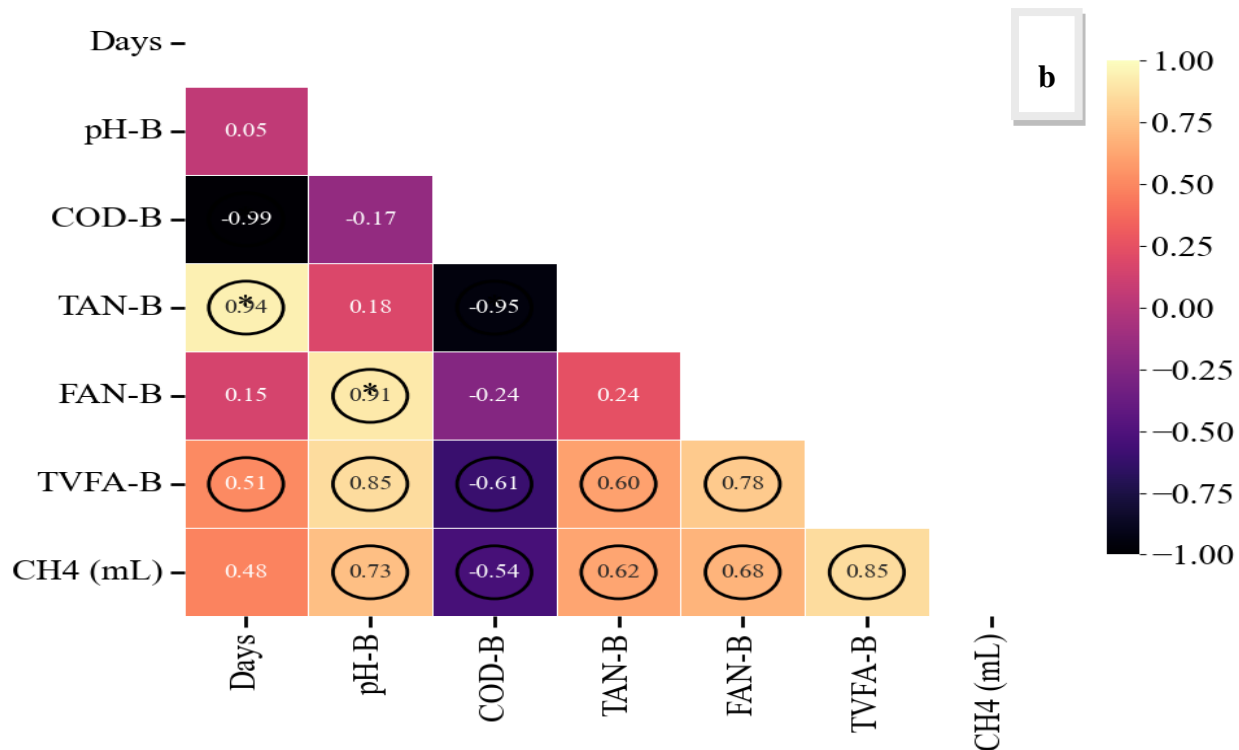
Pearson Pairwise Correlation Heatmap of Reactor-C

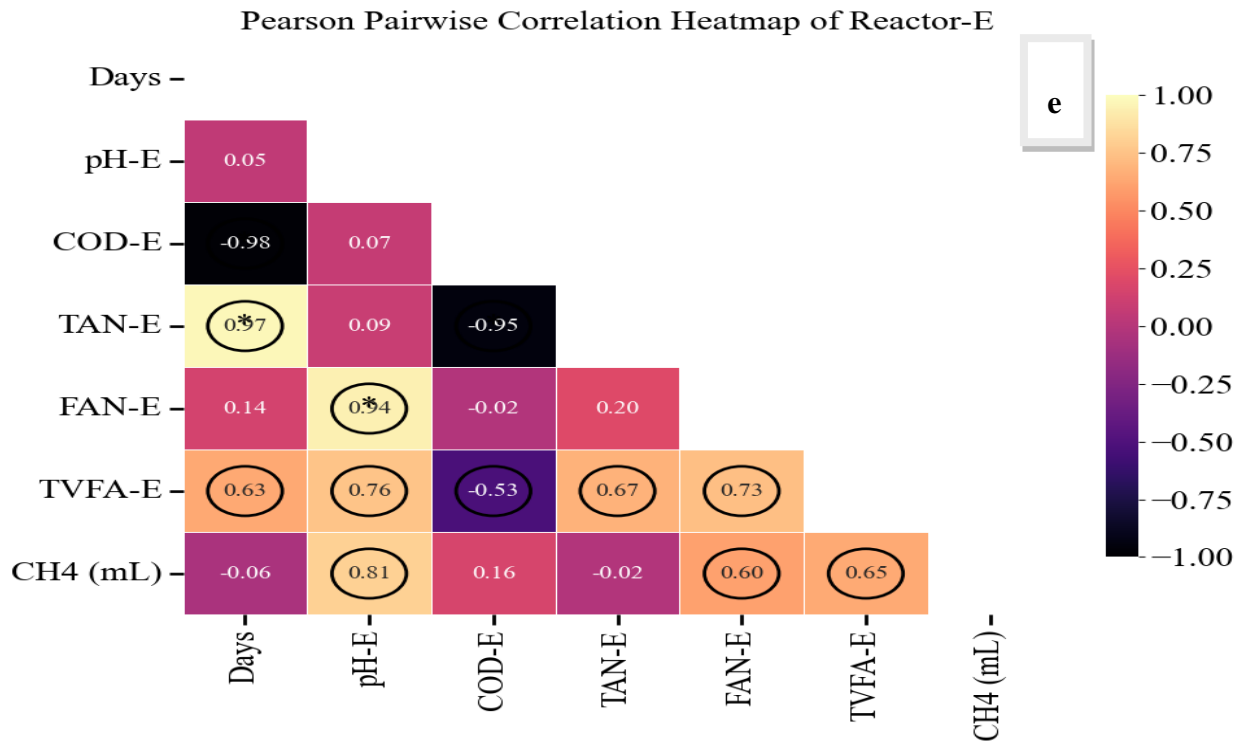Pearson Pairwise Correlation Heatmap of Reactor-D

Figure 4.2: (a) Pairwise Pearson Correlation metrics of Reactors A (b) Pairwise Pearson Correlation metrics of Reactors B (c) Pairwise Pearson Correlation metrics of Reactors C (d) Pairwise Pearson Correlation metrics of Reactors D (e) Pairwise Pearson Correlation metrics of Reactors E

However significant methane production is expected to occur in methanogenesis so, the correlation of the days of digestion and the volume of methane production is lying in the range of 0.34-0.45 shows a positive correlation with 48% of the variance. The correlation metrics provide a deeper understanding of the association between input and target variables, facilitating the interpretation of predictive models [55]. The positive correlation between the days of digestion and methane production reflects the sequential stages of AD where complex substrates are converted into VFAs acetic acid, CO_2 and H_2 which are the precursors of methane production. The correlation metrics connect raw data with model predictions, providing insights into the correlations between input parameters and target variables [59]. The correlation coefficient of 0.60 with methane gas in reactor A, while showing 0.86, 0.57, 0.53 and 0.65 for reactor B-E expediting the correlation of TVFA and methane gas by the above-explained mechanism in four stages explaining 53% to 86% variance the subsequent phase involves the methanogenic archaea. As the

digestion process advances, the population of these microorganisms becomes established and optimized for the breakdown of complex organic matter. The effect of TAN on methane production is inversely related to the correlation coefficients of reactors D and E, The highest methane production is also observed in these two reactors [60]. The most important of all the features is the pH as it decides the fate of the reaction mechanism from reactor A to reactor E explained correlation is 0.33, 0.73, 0.79, 0.75, and 0.81 respectively; the Reactor with the pH 0.75 has the highest performance for methane production expediting pH as a crucial variable in terms of anaerobic treatment process [61].

4.4. Importance of different anaerobic digestion variables: Feature Engineering

Feature selection is an important stage while designing a model where the objective is to find and keep the most relevant variables for prediction explained in Figure 4.3. In the feature importance plot of AD across five reactors with varying Carbon-to-Nitrogen (C/N) ratios ranging from 15-27. Random Forest analysis is used to describe the principle features driving system performance. TVFA was the most influential factor with a substantial importance score of 35% and 71 % for reactor B when C/N was optimized at 18, and another input feature is the organic loading rate (OLR) [62]. The initial low loading rate is likely to be characterized by a period of adaptation for the microbial community. CH_4 production at this stage is not high an increased OLR indicates a larger intake of the organic substrate into the reactor, which means more organic substrate for the microbial community as the OLR increase the methane production is also enhanced since there is more volatile fatty acid accumulation for acetic acid conversion and thus more biogas will generate but at very higher OLR there is a risk of process instability as high loading will lead the microbes to boost their digestion activity and there will be the possibility that the multistage mechanism will disturb and pH shifts its trends to the acidic environment and all the stages cease leaving the process incomplete, so OLR affects methane synthesis, but increasing OLR will increase the methane production to an extent; after that more rise in OLR would not be suitable for microbial digestibility and methane production. Reactors C, D and E explain feature score analysis where the most important feature is pH. [50].
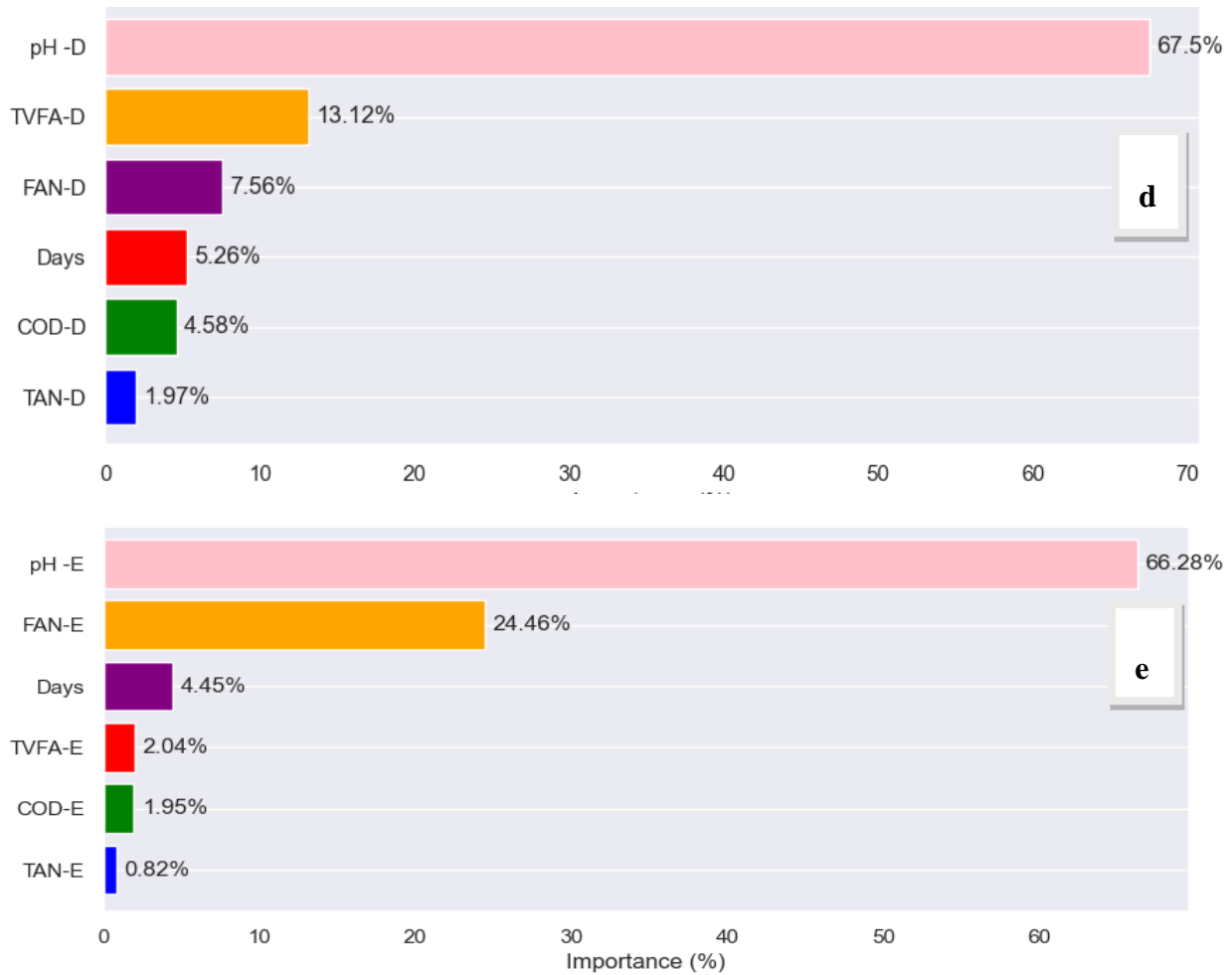
Feature Importance
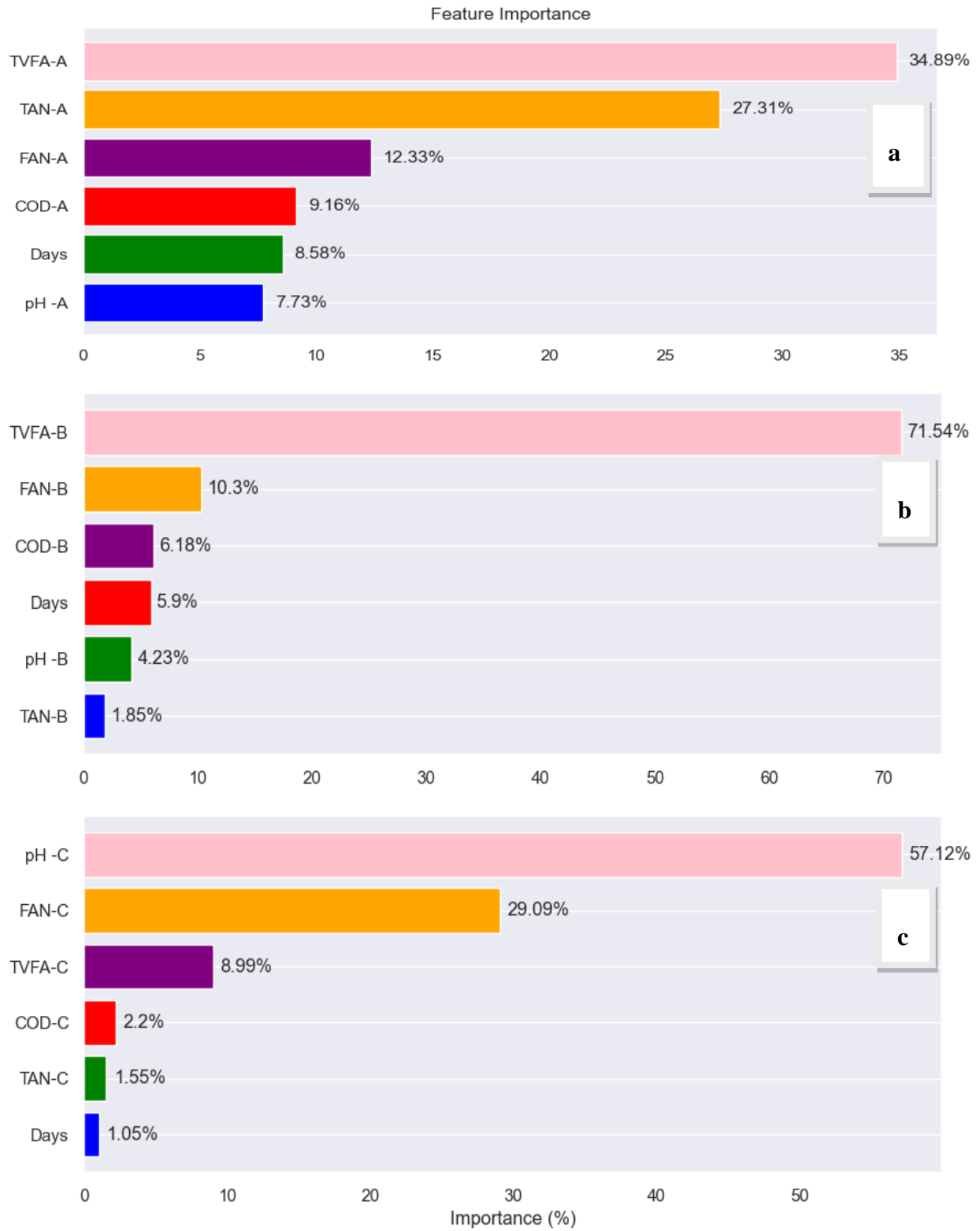
Figure 4.3: (a) Feature Importance via Random Forest Feature Engineering for Reactor A (b) Feature Importance via Random Forest Feature Engineering for Reactor B (c) Feature Importance via Random Forest Feature Engineering for Reactor C (d) Feature Importance via Random Forest Feature Engineering for Reactor D (e) Feature Importance via Random Forest Feature Engineering for Reactor D

As the C/N is increased by 3 digits in all reactors from 15:1 – 27:1 up to 24:1 C/N that was in reactor D methane production increased and then decreased in reactor E hence proving the instability of the AD by adding more substrate after a certain limit process would be retarded as the methane production is ceased [29,63].

4.5 Anaerobic Digestion Modelling Performance of all CSTRs

Table 4.2 discusses studies that have demonstrated the utility of ML algorithms in forecasting outcomes such as methane production, process stability, and substrate degradation. A common theme in these studies is the diverse selection of input features ranging from chemical compositions (C, H, O, N), operational parameters (TS, VS, pH, temperature), and specific biological markers like volatile fatty acids (VFA) and total ammonia nitrogen (TAN). well, each model fits the data and generalizes to new, unseen data across different evaluation stages ensuring a robust understanding.

Table 4.2: A Comparison of the literature review of feature sets and ML algorithms with the present study

Parameters	Algorithms	R^2	Reference
FA, Oil C, TA, TN, TS, VFA/TA, COD, VFA	RFR, LR, ANN, SVM, XGBR	$R^2 = 0.74$	[64]
SMY, Cel, Hem, Lig, SS, CP, C, H, O, N, C/N	ANN, SVM, DTR, KNN,	$R^2 = 0.75\text{-}0.33$	[65]
TS, VS, T, OLR, HRT, VFA, TAN, C/N, CH$_4$ (%), pH	SVM, DTR, KNN, GBR	$R^2 = 0.91\text{-}0.71$	[66]
C, H, O, N, S, Temp	ANN, SVR, RFR, KNN	$R^2 = 0.92$	[67]
TS, VS, VFA, Alkalinity Temp, pH, CO$_2$, CH$_4$	RFR, ANN, KNN, SVR XGBR	$R^2 = 0.92\text{-}0.83$	[68]
Days, COD, pH, TAN, FAN, TVFA (A)	PR, RR, CBR, KNN	$R^2 = 0.72\text{-}0.60$	Present study
Days, COD, pH, TAN, FAN, TVFA (B)	PR, GBR, CBR, KNN	$R^2 = 0.92\text{-}0.86$	Present study
Days, COD, pH, TAN, FAN, TVFA (C)	RFR, XGBR CBR, GBR	$R^2 = 0.95\text{-}0.93$	Present study
Days, COD, pH, TAN, FAN, TVFA (D)	RFR, XGB, CBR, RF	$R^2 = 0.95\text{-}0.73$	Present study
Days, COD, pH, TAN, FAN, TVFA (E)	RFR, XGB, CBR, GBR	$R^2 = 0.94\text{-}0.93$	Present study

These parameters serve as critical indicators of process dynamics within AD systems, and their selection is crucial for developing robust predictive models. For instance, research conducted by [64] applied a combination of multiple ML algorithms including RFR, ANN and LR on the input features including fatty acids, oil contents, total solids and chemical oxygen demand with an optimal performing R^2 value of 0.74. Similar other studies [65,69] have employed SVR, KNNR and DTR on various AD operational parameters with R^2 ranging from 0.33-0.75 targeting the complexity of the set of features and modelling approach. A notable study by [66] integrated advanced gradient boosting methods, such as the Gradient Boosting Regressor (GBR) and Extreme Gradient Boosting (XGBR), alongside SVM and KNN to model outputs like methane

concentration and process stability based on a multi-variate input of TS, VS, VFA, pH, and temperature. This study achieved an impressive R^2 range of 0.91 to 0.71, showcasing the effectiveness of boosting algorithms in capturing non-linear relationships in AD processes. More recent works, such as that of [67], have expanded the feature sets to include elemental compositions (C, H, O, N) alongside temperature and operational metrics like organic loading rate (OLR) and hydraulic retention time (HRT). These studies utilized hybrid modelling approaches, combining ANN, Support Vector Regressor (SVR), and RFR, and achieved an R^2 value as high as 0.92. The focus on hybrid models and ensemble techniques in this area reflects the increasing trend toward maximizing model accuracy by leveraging the strengths of multiple algorithms. In the context of the present study, a combination of advanced ML techniques, including Polynomial Regression (PR), Ridge Regression (RR), and case-based reasoning (CBR), alongside tree-based

Table 4.3: Evaluation of determination coefficient R^2 of Reactors A, B, C, D and E datasets

CSTRs	Regression Model	Algorithms	Training R^2	Testing R^2	Cross Validation R^2
Reactor-A	Linear	PR	0.773	0.609	0.632
		RR	0.648	0.605	0.758
	Ensemble	CBR	0.859	0.742	0.676
	Non-Linear	KNNR	0.688	0.722	0.709
Reactor-B	Linear	PR	0.969	0.919	0.888
	Ensemble	GBR	0.998	0.836	0.914
		CBR	0.978	0.895	0.850
	Non-Linear	KNNR	0.937	0.924	0.846
Reactor-C	Ensemble	RFR	0.992	0.940	0.910
		GBR	0.999	0.968	0.914
		XGBR	0.999	0.956	0.886
		CBR	0.979	0.958	0.897
Reactor-D	Ensemble	RFR	0.991	0.951	0.963
		XGBR	0.999	0.964	0.843
		CBR	0.976	0.852	0.935
	Non-Linear	KNNR	0.924	0.739	0.875
Reactor-E	Ensemble	RFR	0.990	0.949	0.813
		GBR	0.995	0.940	0.849
		XGBR	0.998	0.942	0.771
		CBR	0.975	0.935	0.818

methods like RFR and boosting methods like XGB and GBR, was applied to predict key AD process indicators such as chemical oxygen demand (COD), pH, total volatile fatty acids (TVFA), and nitrogen compounds (TAN, FAN). The comparison of results of current studies in Table 4.3 includes the R^2 values for training, testing, and cross-validation, which collectively indicate how

Linear Regression Methods including PR and RR, generally performed well but were outperformed by more sophisticated models, especially when dealing with more complex data. Their R^2 scores in testing ranged from moderate to high, indicating that while they can capture basic trends, they may struggle with intricate patterns in the data. The only Non-Linear Regression Method applied was the KNNR, a representative of non-linear methods, which consistently provided competitive R^2 scores [66], particularly in Reactors A and B. This model's strength lies in its ability to model data without assuming a linear relationship, making it versatile for capturing more complex interactions. Ensemble Regression Methods including RFR [64,68] GBR, XGBR [68,70] and CBR generally delivered the highest R^2 scores across all reactors. These ensemble learning models capture complex, non-linear relationships and reduce overfitting, making them ideal for datasets where intricate patterns drive the output. [65]. The success of Reactor D's ensemble models, particularly XGBR, underscores their effectiveness in achieving high methane yields. Reactor D stands out as the best-performing dataset for methane yield, as evidenced by the R^2 results. Figure 4.4-4.8 demonstrates the goodness of fit of each ML algorithm for the comparative analysis. The Ensemble learning models, particularly the XGBR, achieved the highest R^2 scores with 0.999 on training, 0.964 on testing, and a robust 0.843 on cross-validation. This indicates an incomparable fit during training and strong generalization during the testing and cross-validation phases. The RFR also performed remarkably well, with R^2 values of 0.991 training, 0.951 testing, and 0.963 cross-validation. These high scores across all phases underscore the model's ability to capture complex patterns in the data, contributing to its effectiveness in predicting methane yield. Additionally, CBR, another Ensemble method, showed solid performance with an R^2 of 0.976, 0.852 and 0.935 training, testing and cross-validation respectively highlighting its reliability and accuracy. In reactor A the linear regression models PR and RR exhibited moderate performance, with R^2 values around 0.6-0.7 during testing, indicating a decent but not exceptional model fit their capacity to generalize the complex linear patterns was limited. In the Ensemble method, CBR performed better with a testing R^2 of 0.742, showing it could model intricate relationships in the data. The non-linear model KNNR also provided competitive results with a testing R^2 of 0.722, demonstrating its capability to capture non-linear patterns.
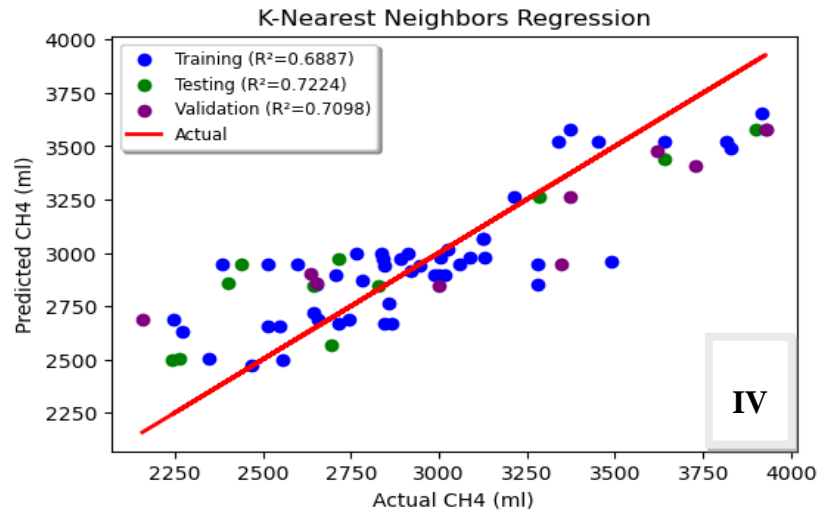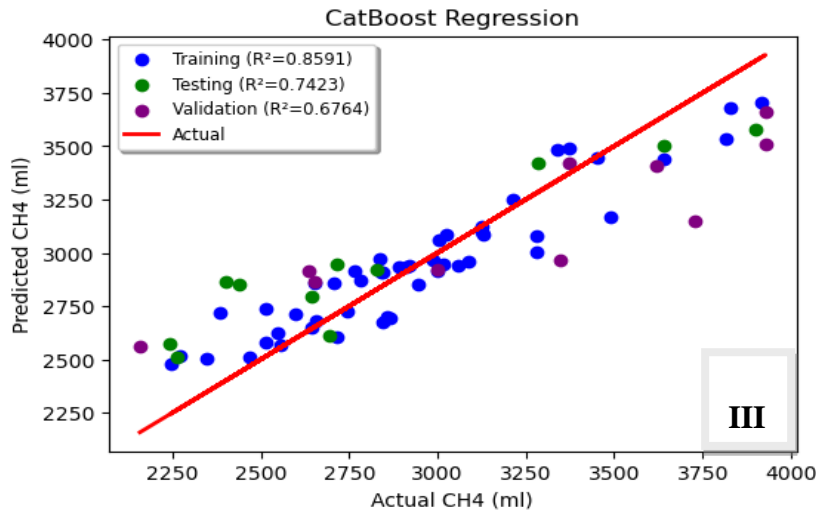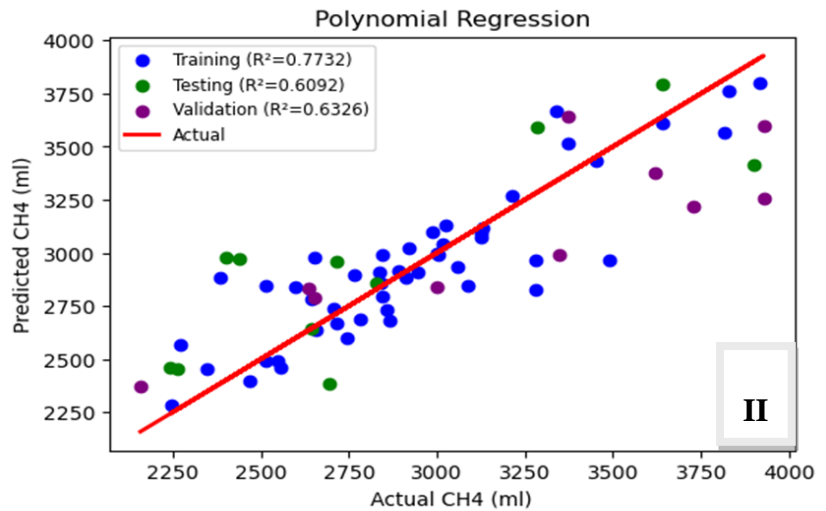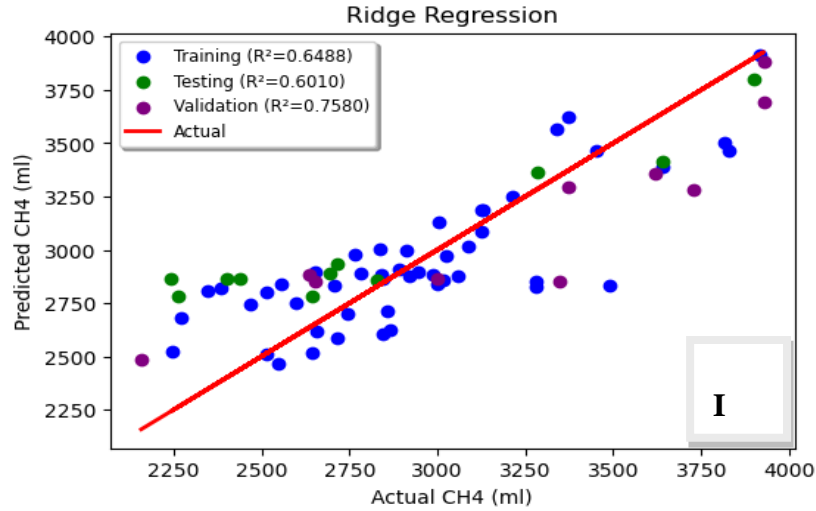
Figure 4.4: Reactor A; Actual vs Predicted CH$_4$ (ml) Plots of Regression Models; I: Ridge II: Polynomial III: CatBoost IV: k-Nearest Neighbor
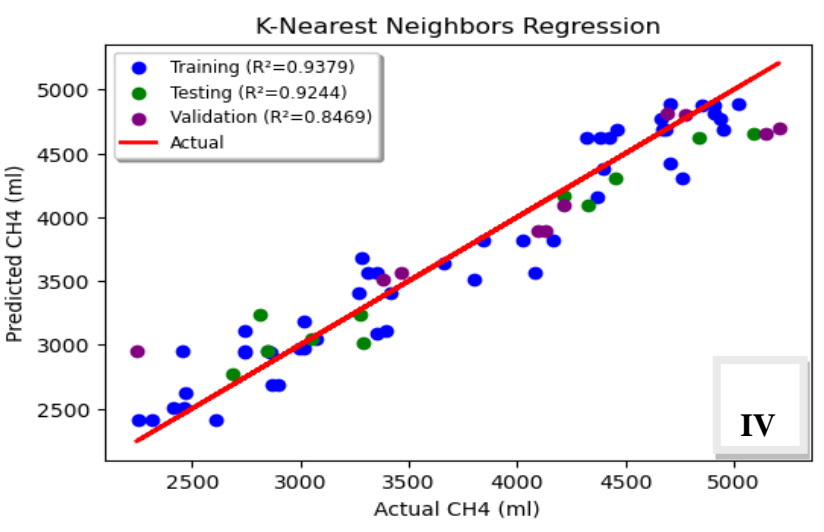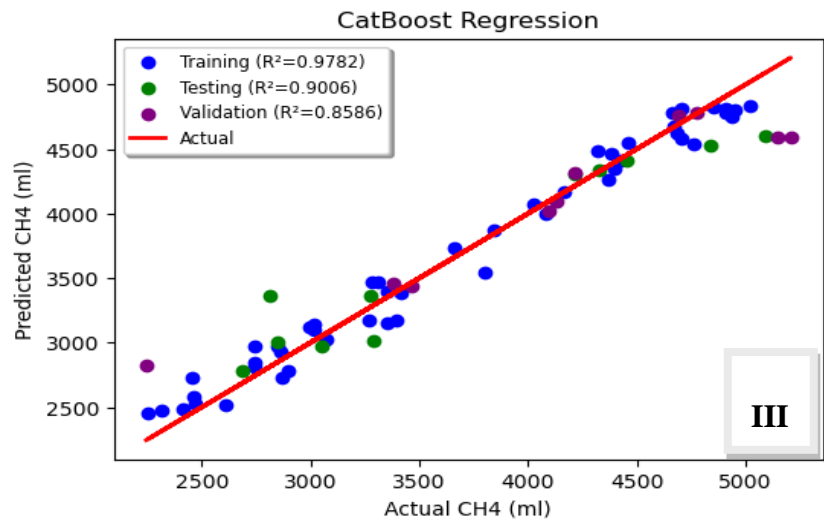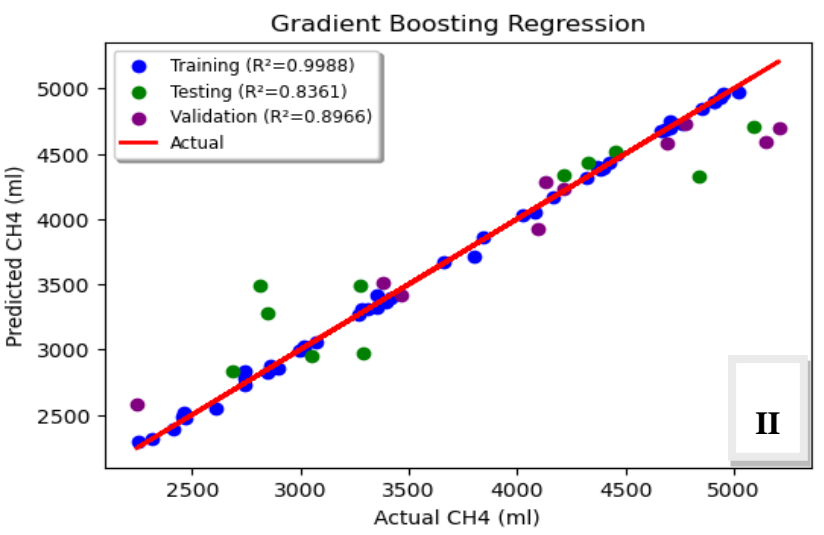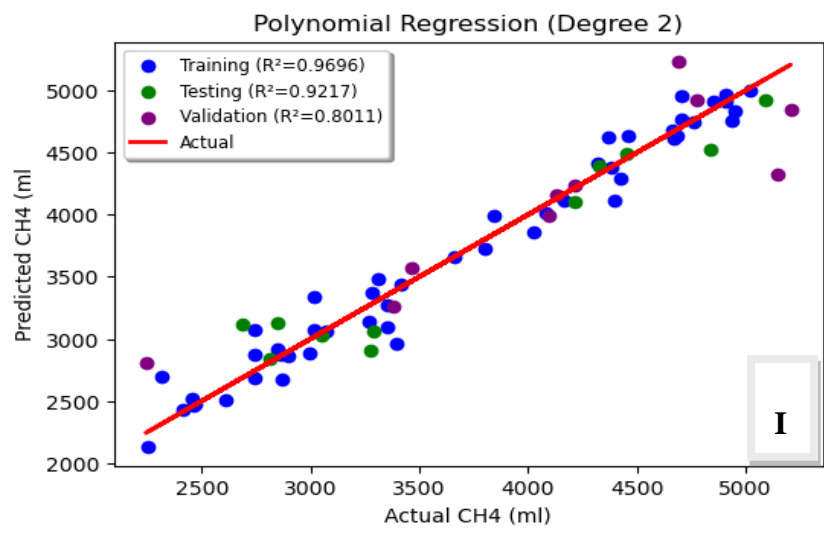
Figure 4.5: Reactor B; Actual vs Predicted CH$_4$ (ml) Plots of Regression Models; I: Polynomial II: Gradient Boosting III: CatBoost IV: k-Nearest Neighbor
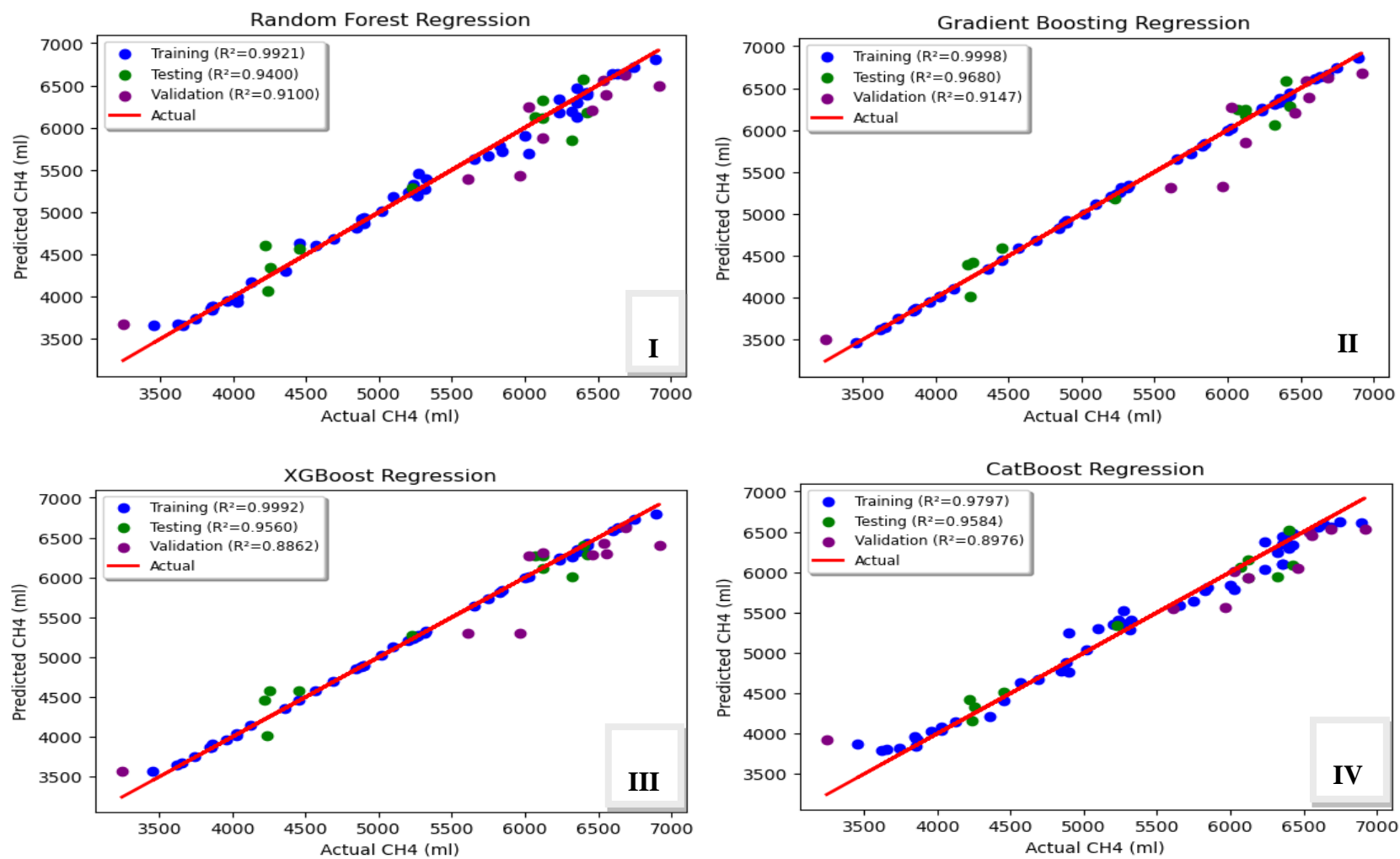
Figure 4.6: Reactor C; Actual vs Predicted CH$_4$ (ml) Plots of Regression Models; I: Random Forest II: Gradient Boosting III: XGBoost IV: CatBoost
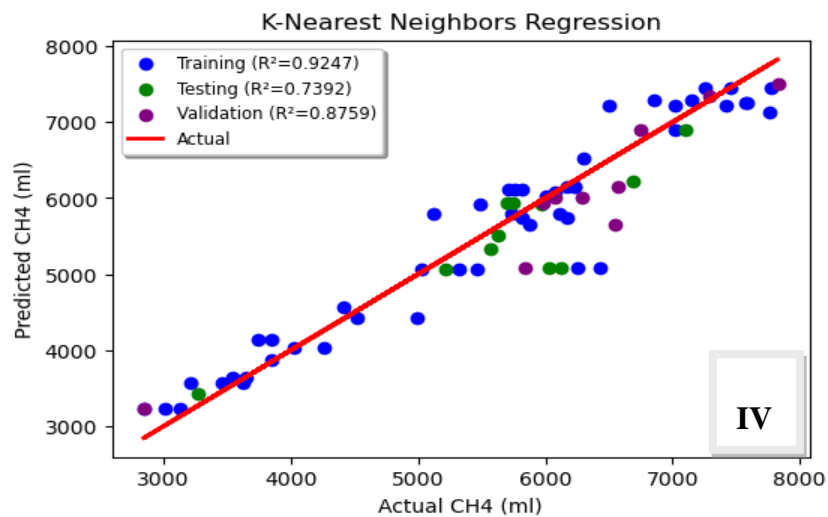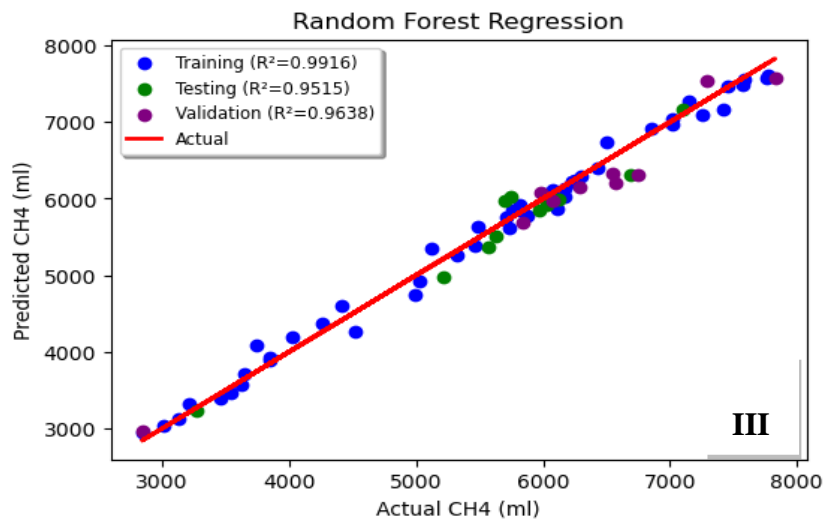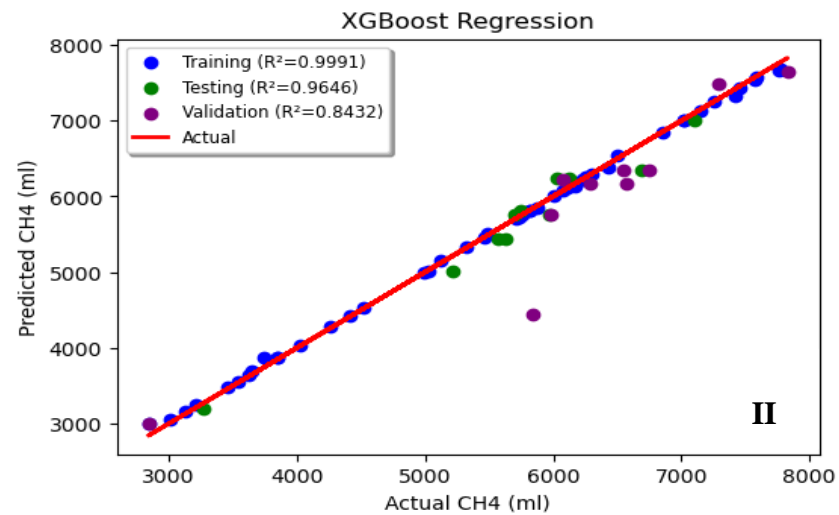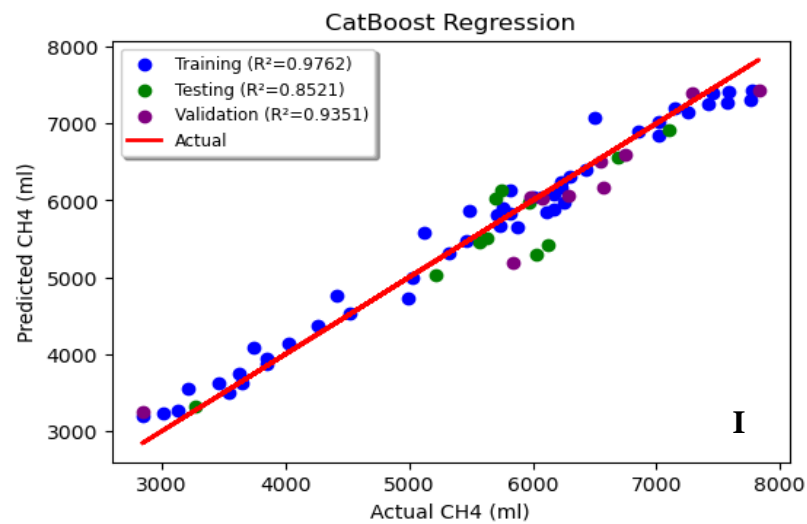
Figure 4.7: Reactor D; Actual vs Predicted CH$_4$ (ml) Plots of Regression Models; I: CatBoost II: XGBoost III: Random Forest IV: k-Nearest Neighbor: CatBoost
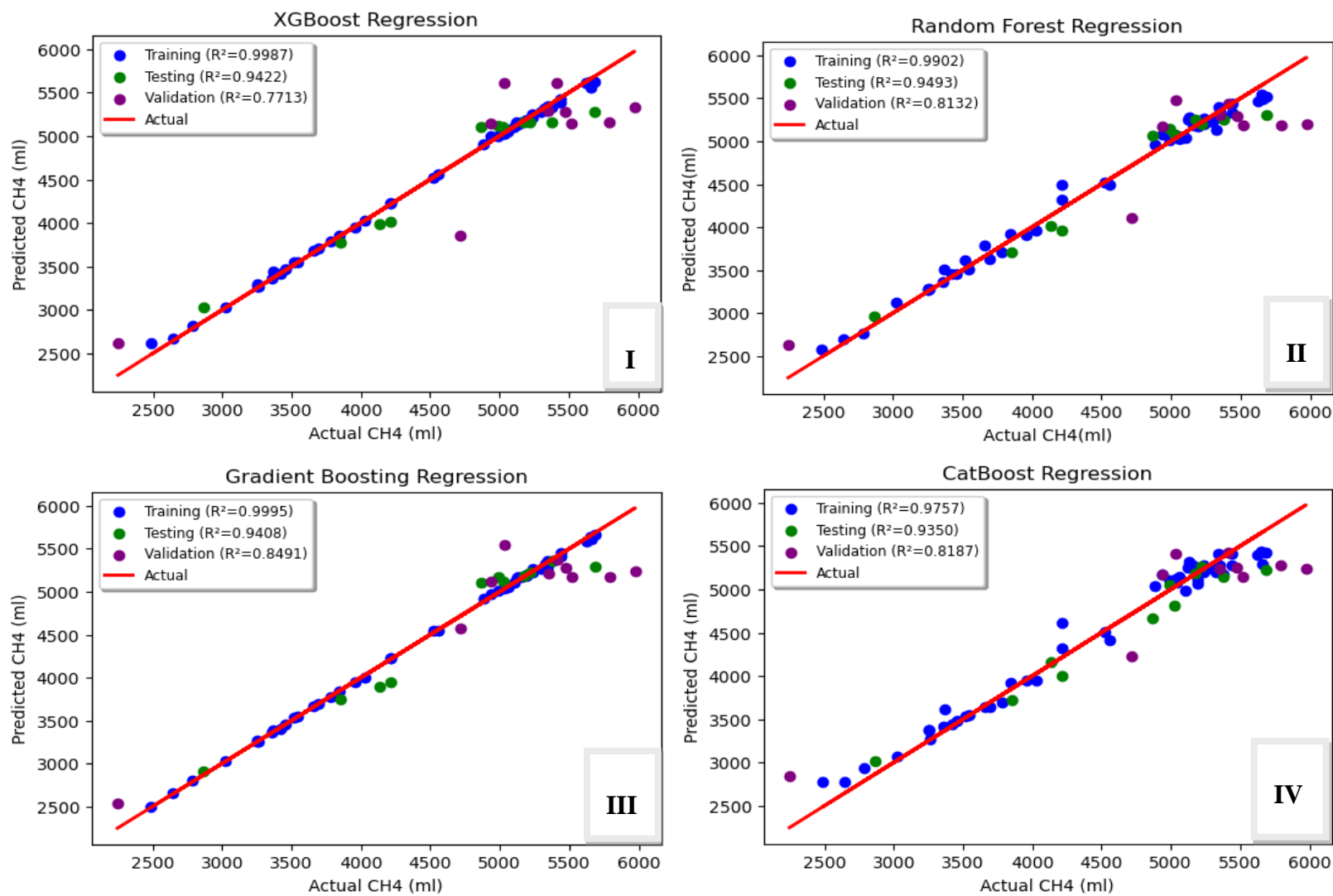
Figure 4.8: Reactor E; Actual vs Predicted CH$_4$ (ml) Plots of Regression Models; I: XGBoost II: Random Forest III: Gradient Boosting IV: CatBoost

In reactor B the PR model excelled with an R² of 0.919 in testing, showing strong performance as a Linear model. However, Ensemble methods like GBR and CBR also delivered high R² scores (0.836 and 0.895, respectively) during testing, these results highlighting their robustness to generalize well. The KNNR further affirmed its efficacy as a Non-Linear model with a testing R² of 0.924. In reactor C the ensemble methods emerged as a dominant approach, with GBR achieving the uppermost testing R² of 0.968, followed closely by XGBR and CBR with R² scores of 0.956 and 0.958, respectively. The strong performance across these models suggests that Reactor C's data benefits significantly from complex, ensemble-based approaches. In reactor E Ensemble models again demonstrated to be the superior performers, with RFR, GBR, and XGBR all achieving testing R² values around 0.94-0.95. This consistency across multiple Ensemble methods indicates their robustness and adaptability in predicting outcomes further reinforcing their applicability to complex real-world datasets.

CHAPTER 5: CONCLUSION AND FUTURE RECOMMENDATION

5.1 Conclusion

The study illustrates that optimizing the Carbon to Nitrogen ratio (C/N) 15:1-27:1 and the Organic Loading Rate (OLR) to around 1.5-7.5 g VS/L/d in Continuous Stirred Tank Reactors (CSTRs) significantly enhances methane production during AD of chicken manure and rice straw, achieving yields of 126.50% methane in C/N 24:1. This optimization supports microbial activity, enhancing the breakdown of organic material and ensuring process stability. Through the application of ensemble learning regression models, methane yields were predicted with remarkable precision, achieving R^2 values of 0.92 and 0.89. These models effectively captured the intricate relationships between variables such as the C/N ratio, OLR, and methane production, demonstrating the power of advanced predictive techniques. The integration of ML not only validates the optimization strategy but also provides a robust predictive framework, offering a scalable solution for improving biogas production efficiency and contributing to sustainable energy goals of cleaner and cheaper energy as well as clean water and sanitation.

5.2 Future Perspective

Integrating ML with IoT (Internet of Things) devices such as temperature sensors, pH sensors, and biogas analyzers can revolutionize real-time experimental monitoring and control in biogas production [71]; These IoT devices collect the operational data continuously which will be transmitted to a central processing unit for analysis. If certain operational parameters such as temperature deviate from their range the trained ML algorithm from the previous dataset identifies the incoming data pattern; and analyzes anomalous behaviour such as unexpected rise or fall in temperature and its potential impact on AD. Immediate action such as adjusting the thermostat and makes to nullify the distortion in the process restores the optimal conditions. This automated system loop

ensures the process remains stable and efficient by reverting the change accordingly. Additionally, the extensive data collection by this developed system can be used to employ advanced ML techniques such as neural networks. This synergy of IoT and ML enables predictive modelling analytics, and automated optimization of the AD mechanism leading to more sustainable AD processes for biogas production this synergy will drive more significant enhancements in the field.

REFERENCES

[1]   A. Bakhtiar, A. Aslani, S.M. Hosseini, Challenges of diffusion and commercialization of bioenergy in developing countries, Renew. Energy 145 (2020). https://doi.org/10.1016/j.renene.2019.06.126.

[2]   F. Dell'Anna, Green jobs and energy efficiency as strategies for economic growth and the reduction of environmental impacts, Energy Policy 149 (2021). https://doi.org/10.1016/j.enpol.2020.112031.

[3]   M.N.V. Prasad, Recovery of Resources From Biowaste for Pollution Prevention, in: Environ. Mater. Waste Resour. Recover. Pollut. Prev., 2016. https://doi.org/10.1016/B978-0-12-803837-6.00001-9.

[4]   M. Batool, L. Shahzad, An analytical study on municipal wastewater to energy generation, current trends, and future prospects in South Asian developing countries (an update on Pakistan scenario), Environ. Sci. Pollut. Res. 28 (2021). https://doi.org/10.1007/s11356-021-14029-8.

[5]   Tasnim, F., S.A. Iqbal, A.R. Chowdhury, Biogas production from anaerobic codigestion of cow manure with kitchen waste and Water Hyacinth., Renew. Energy 109 (2017) 434–439.

[6]   W. Li, F. Chien, Q.T. Ngo, T.D. Nguyen, S. Iqbal, A.R. Bilal, Vertical financial disparity, energy prices and emission reduction: Empirical insights from Pakistan, J. Environ. Manage. 294 (2021) 112946. https://doi.org/10.1016/j.jenvman.2021.112946.

[7]   S. Khan, A. Nisar, B. Wu, Q.L. Zhu, Y.W. Wang, G.Q. Hu, M. xiong He, Bioenergy production in Pakistan: Potential, progress, and prospect, Sci. Total Environ. 814 (2022). https://doi.org/10.1016/j.scitotenv.2021.152872.

[8]   Gielen, D., F. Boshell, D. Saygin, M.D. Bazilian, N. Wagner, R. Gorini, The role of renewable energy in the global energy transformation., Energy Strateg. Rev. 24 (2019) 38–50.

[9]   J. Lee, D. Choi, Y.S. Ok, S.R. Lee, E.E. Kwon, Enhancement of energy recovery from chicken manure by pyrolysis in carbon dioxide, J. Clean. Prod. 164 (2017). https://doi.org/10.1016/j.jclepro.2017.06.217.

[10]   L. Jurgutis, A. Slepetiene, J. Volungevicius, K. Amaleviciute-Volunge, Biogas production from chicken manure at different organic loading rates in a mesophilic full scale anaerobic digestion plant, Biomass and Bioenergy 141 (2020). https://doi.org/10.1016/j.biombioe.2020.105693.

[11]   M. Wakeel, T. Hayat, N.S. Shah, J. Iqbal, Z.U. Haq Khan, G.M. Shah, A. Rasool,

Biogas Energy Resources in Pakistan Status, Potential, and Barriers, Util. Policy 84 (2023) 101643. https://doi.org/10.1016/j.jup.2023.101643.

[12]   P. Division, Yearbook 2022-23 Ministry of Energy (Power Division) 1, (2022) 1–72.

[13]   T. Akmal, F. Jamil, Assessing health damages from improper disposal of solid waste in metropolitan islamabad–rawalpindi, Pakistan, Sustain. 13 (2021). https://doi.org/10.3390/su13052717.

[14]   W. Zhou, X. Chen, Y. Wang, N. Tuersun, M. Ismail, C. Cheng, Z. Li, Q. Song, Y. Wang, C. Ma, Anaerobic co-digestion of textile dyeing sludge: Digestion efficiency and heavy metal stability, Sci. Total Environ. 801 (2021). https://doi.org/10.1016/j.scitotenv.2021.149722.

[15]   W. Khan, K. Shehzada, T. Bibi, S. Ul Islam, S. Wali Khan, Performance evaluation of Khyber Paktunkhwa Rice Husk Ash (RHA) in improving mechanical behavior of cement, Constr. Build. Mater. 176 (2018). https://doi.org/10.1016/j.conbuildmat.2018.04.213.

[16]   T.A. Memon, Assessment of Rice Residues as Potential Energy Source in Pakistan, Sukkur IBA J. Emerg. Technol. 5 (2022). https://doi.org/10.30537/sjet.v5i1.982.

[17]   A. Ali, R.B. Mahar, S. Panhwar, H.A. Keerio, N.H. Khokhar, F. Suja, L. Rundong, Generation of Green Renewable Energy Through Anaerobic Digestion Technology (ADT): Technical Insights Review, Waste and Biomass Valorization 14 (2023). https://doi.org/10.1007/s12649-022-02001-7.

[18]   P.M.V. Subbarao, T.C. D' Silva, K. Adlak, S. Kumar, R. Chandra, V.K. Vijay, Anaerobic digestion as a sustainable technology for efficiently utilizing biomass in the context of carbon neutrality and circular economy, Environ. Res. 234 (2023). https://doi.org/10.1016/j.envres.2023.116286.

[19]   R. Cano, A. Nielfa, M. Fdz-Polanco, Thermal hydrolysis integration in the anaerobic digestion process of different solid wastes: Energy and economic feasibility study, Bioresour. Technol. 168 (2014). https://doi.org/10.1016/j.biortech.2014.02.007.

[20]   A. Hornung, Transformation of Biomass: Theory to Practice, 2014. https://doi.org/10.1002/9781118693643.

[21]   M.J. McInerney, M.P. Bryant, Basic Principles of Bioconversions in Anaerobic Digestion and Methanogenesis, in: Biomass Convers. Process. Energy Fuels, 1981. https://doi.org/10.1007/978-1-4757-0301-6_15.

[22]   I. Andrade Cruz, W. Chuenchart, F. Long, K.C. Surendra, L. Renata Santos Andrade, M. Bilal, H. Liu, R. Tavares Figueiredo, S.K. Khanal, L. Fernando Romanholo Ferreira, Application of machine learning in anaerobic digestion: Perspectives and challenges, Bioresour. Technol. 345 (2022).

https://doi.org/10.1016/j.biortech.2021.126433.

[23] A. Saravanan, P. Senthil Kumar, G. Rangasamy, R. Hariharan, R. V. Hemavathy, P.D. Deepika, K. Anand, S. Karthika, Strategies for enhancing the efficacy of anaerobic digestion of food industry wastewater: An insight into bioreactor types, challenges, and future scope, Chemosphere 310 (2023) 136856. https://doi.org/10.1016/j.chemosphere.2022.136856.

[24] A. Raheem, M.Y. Hassan, R. Shakoor, Bioenergy from anaerobic digestion in Pakistan: Potential, development and prospects, Renew. Sustain. Energy Rev. 59 (2016). https://doi.org/10.1016/j.rser.2016.01.010.

[25] Y. Zhang, Y. Feng, Z. Ren, R. Zuo, T. Zhang, Y. Li, Y. Wang, Z. Liu, Z. Sun, Y. Han, L. Feng, M. Aghbashlo, M. Tabatabaei, J. Pan, Tree-based machine learning model for visualizing complex relationships between biochar properties and anaerobic digestion, Bioresour. Technol. 374 (2023) 128746. https://doi.org/10.1016/j.biortech.2023.128746.

[26] A. Sonwai, P. Pholchan, N. Tippayawong, Machine learning approach for determining and optimizing influential factors of biogas production from lignocellulosic biomass, Bioresour. Technol. 383 (2023) 129235. https://doi.org/10.1016/j.biortech.2023.129235.

[27] G. Ali, M.K. Bashir, H. Ali, M.H. Bashir, Utilization of rice husk and poultry wastes for renewable energy potential in Pakistan: An economic perspective, Renew. Sustain. Energy Rev. 61 (2016). https://doi.org/10.1016/j.rser.2016.03.014.

[28] E. Rossi, I. Pecorini, R. Iannelli, Multilinear Regression Model for Biogas Production Prediction from Dry Anaerobic Digestion of OFMSW, Sustain. 14 (2022). https://doi.org/10.3390/su14084393.

[29] M. Hassan, W. Ding, M. Umar, G. Rasool, Batch and semi-continuous anaerobic co-digestion of goose manure with alkali solubilized wheat straw: A case of carbon to nitrogen ratio and organic loading rate regression optimization, Bioresour. Technol. 230 (2017) 24–32. https://doi.org/10.1016/j.biortech.2017.01.025.

[30] I. TG, I. Haq, A.S. Kalamdhad, 14 - Factors affecting anaerobic digestion for biogas production: a review, Adv. Org. Waste Manag. (2022).

[31] E. Bellodi, R. Zese, F. Riguzzi, E. Lamma, Introduction to Machine Learning, in: Mach. Learn. Non-Volatile Memories, 2022. https://doi.org/10.1007/978-3-031-03841-9_1.

[32] V. Kanade, What is machine learning? Understanding types & applications, Spiceworks (2022).

[33] T.H. Tsui, M.C.M. van Loosdrecht, Y. Dai, Y.W. Tong, Machine learning and circular bioeconomy: Building new resource efficiency from diverse waste streams,

Bioresour. Technol. 369 (2023). https://doi.org/10.1016/j.biortech.2022.128445.

[34] X. Dong, Z. Yu, W. Cao, Y. Shi, Q. Ma, A survey on ensemble learning, Front. Comput. Sci. 14 (2020). https://doi.org/10.1007/s11704-019-8208-z.

[35] I.D. Mienye, Y. Sun, A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects, IEEE Access 10 (2022). https://doi.org/10.1109/ACCESS.2022.3207287.

[36] D. Borup, B.J. Christensen, N.S. Mühlbach, M.S. Nielsen, Targeting predictors in random forest regression, Int. J. Forecast. 39 (2023) 841–868. https://doi.org/10.1016/j.ijforecast.2022.02.010.

[37] N. Gunasekara, B. Pfahringer, H. Gomes, A. Bifet, Gradient boosted trees for evolving data streams, Mach. Learn. 113 (2024) 3325–3352. https://doi.org/10.1007/s10994-024-06517-y.

[38] Z. Arif Ali, Z. H. Abduljabbar, H. A. Tahir, A. Bibo Sallow, S.M. Almufti, eXtreme Gradient Boosting Algorithm with Machine Learning: a Review, Acad. J. Nawroz Univ. 12 (2023) 320–334. https://doi.org/10.25007/ajnu.v12n2a1612.

[39] A. Jadbabaie, A. Makur, D. Shah, Gradient-Based Empirical Risk Minimization Using Local Polynomial Regression, Stoch. Syst. 5238 (2024) 1–40. https://doi.org/10.1287/stsy.2022.0003.

[40] H. Kaya, E. Guler, V. Kırmacı, Prediction of temperature separation of a nitrogen-driven vortex tube with linear, kNN, SVM, and RF regression models, Neural Comput. Appl. 35 (2023) 6281–6291. https://doi.org/10.1007/s00521-022-08030-6.

[41] M. Hassan, W. Ding, Z. Shi, S. Zhao, Methane enhancement through co-digestion of chicken manure and thermo-oxidative cleaved wheat straw with waste activated sludge: A C/N optimization case, Bioresour. Technol. 211 (2016) 534–541. https://doi.org/10.1016/j.biortech.2016.03.148.

[42] K. Sahoo, A.K. Samal, J. Pramanik, S.K. Pani, Exploratory data analysis using python, Int. J. Innov. Technol. Explor. Eng. 8 (2019) 4727–4735. https://doi.org/10.35940/ijitee.L3591.1081219.

[43] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, PeerJ Comput. Sci. 7 (2021). https://doi.org/10.7717/PEERJ-CS.623.

[44] Z. Mei, X. Liu, X. Huang, D. Li, Z. Yan, Y. Yuan, Y. Huang, Anaerobic Mesophilic Codigestion of Rice Straw and Chicken Manure: Effects of Organic Loading Rate on Process Stability and Performance, Appl. Biochem. Biotechnol. 179 (2016). https://doi.org/10.1007/s12010-016-2035-6.

[45] Y. Li, R. Zhang, C. Chen, G. Liu, Y. He, X. Liu, Biogas production from co-digestion of corn stover and chicken manure under anaerobic wet, hemi-solid, and solid state conditions, Bioresour. Technol. 149 (2013). https://doi.org/10.1016/j.biortech.2013.09.091.

[46] J. Feng, Y. Li, E. Zhang, J. Zhang, W. Wang, Y. He, G. Liu, C. Chen, Solid-State Co-digestion of NaOH-Pretreated Corn Straw and Chicken Manure Under Mesophilic Condition, Waste and Biomass Valorization 9 (2018). https://doi.org/10.1007/s12649-017-9834-z.

[47] M.A. Rahman, H.B. Møller, C.K. Saha, M.M. Alam, The effect of temperature on the anaerobic co-digestion of poultry droppings and sugar mill press mud, Biofuels 13 (2022). https://doi.org/10.1080/17597269.2019.1649902.

[48] M.A. Rahman, H.B. Møller, C.K. Saha, M.M. Alam, R. Wahid, L. Feng, Anaerobic co-digestion of poultry droppings and briquetted wheat straw at mesophilic and thermophilic conditions: Influence of alkali pretreatment, Renew. Energy 128 (2018). https://doi.org/10.1016/j.renene.2018.05.076.

[49] M.A. Rahman, H.B. Møller, C.K. Saha, M.M. Alam, R. Wahid, L. Feng, Optimal ratio for anaerobic co-digestion of poultry droppings and lignocellulosic-rich substrates for enhanced biogas production, Energy Sustain. Dev. 39 (2017). https://doi.org/10.1016/j.esd.2017.04.004.

[50] H. A. ali, J. J. Faraj, F. M. Hussien, Effect of pH on biogas production during anaerobic digestion, J. Univ. Shanghai Sci. Technol. 23 (2021) 224–231. https://doi.org/10.51201/jusst/21/08369.

[51] M. Hassan, W. Ding, J. Bi, E. Mehryar, Z.A.A. Talha, H. Huang, Methane enhancement through oxidative cleavage and alkali solubilization pre-treatments for corn stover with anaerobic activated sludge, Bioresour. Technol. 200 (2016) 405–412. https://doi.org/10.1016/j.biortech.2015.09.115.

[52] Y. Cai, M. Zhu, X. Meng, J.L. Zhou, H. Zhang, X. Shen, The role of biochar on alleviating ammonia toxicity in anaerobic digestion of nitrogen-rich wastes: A review, Bioresour. Technol. 351 (2022). https://doi.org/10.1016/j.biortech.2022.126924.

[53] S. Greses, E. Tomás-Pejó, C. Gónzalez-Fernández, Agroindustrial waste as a resource for volatile fatty acids production via anaerobic fermentation, Bioresour. Technol. 297 (2020). https://doi.org/10.1016/j.biortech.2019.122486.

[54] S. Wang, D. Li, K. Zhang, Y. Ma, F. Liu, Z. Li, X. Gao, W. Gao, L. Du, Effects of initial volatile fatty acid concentrations on process characteristics, microbial communities, and metabolic pathways on solid-state anaerobic digestion, Bioresour. Technol. 369 (2023). https://doi.org/10.1016/j.biortech.2022.128461.

[55] H.M. Wu, Y.J. Tien, M.R. Ho, H.G. Hwu, W.C. Lin, M.H. Tao, C.H. Chen,

Covariate-adjusted heatmaps for visualizing biological data via correlation decomposition, Bioinformatics 34 (2018). https://doi.org/10.1093/bioinformatics/bty335.

[56]   X. Zheng, L. Xue, A. Qu, Time-varying correlation structure estimation and local-feature detection for spatio-temporal data, J. Multivar. Anal. 168 (2018). https://doi.org/10.1016/j.jmva.2018.07.012.

[57]   M.A. Salam, A.T. Azar, M.S. Elgendy, K.M. Fouad, The Effect of Different Dimensionality Reduction Techniques on Machine Learning Overfitting Problem, Int. J. Adv. Comput. Sci. Appl. 12 (2021) 641–655. https://doi.org/10.14569/IJACSA.2021.0120480.

[58]   A. Bashir, M. Nafees, N.H. Khan, An assessment of Poultry Farms for Poultry Litter Production, Disposal and Its Use as a Fertilizer, Waste Forum (2022).

[59]   D. Wheeler, M. Tiefelsdorf, Multicollinearity and correlation among local regression coefficients in geographically weighted regression, J. Geogr. Syst. 7 (2005) 161–187. https://doi.org/10.1007/s10109-005-0155-6.

[60]   I. Morozova, N. Nikulina, H. Oechsner, J. Krümpel, A. Lemmer, Effects of increasing nitrogen content on process stability and reactor performance in anaerobic digestion, Energies 13 (2020). https://doi.org/10.3390/en13051139.

[61]   Y. Cai, D. Gallegos, Z. Zheng, W. Stinner, X. Wang, J. Pröter, F. Schäfer, Exploring the combined effect of total ammonia nitrogen, pH and temperature on anaerobic digestion of chicken manure using response surface methodology and two kinetic models, Bioresour. Technol. 337 (2021). https://doi.org/10.1016/j.biortech.2021.125328.

[62]   J. Jiang, S. He, X. Kang, Y. Sun, Z. Yuan, T. Xing, Y. Guo, L. Li, Effect of Organic Loading Rate and Temperature on the Anaerobic Digestion of Municipal Solid Waste: Process Performance and Energy Recovery, Front. Energy Res. 8 (2020). https://doi.org/10.3389/fenrg.2020.00089.

[63]   Q. Yu, L. Feng, X. Zhen, Effects of organic loading rate and temperature fluctuation on the microbial community and performance of anaerobic digestion of food waste, Environ. Sci. Pollut. Res. 28 (2021). https://doi.org/10.1007/s11356-020-11548-8.

[64]   C. Li, P. He, W. Peng, F. Lü, R. Du, H. Zhang, Exploring available input variables for machine learning models to predict biogas production in industrial-scale biogas plants treating food waste, J. Clean. Prod. 380 (2022) 135074. https://doi.org/10.1016/j.jclepro.2022.135074.

[65]   Z. Wang, X. Peng, A. Xia, A.A. Shah, H. Yan, Y. Huang, X. Zhu, X. Zhu, Q. Liao, Comparison of machine learning methods for predicting the methane production from anaerobic digestion of lignocellulosic biomass, Energy 263 (2023) 125883. https://doi.org/10.1016/j.energy.2022.125883.

[66] P. Ganeshan, A. Bose, J. Lee, S. Barathi, K. Rajendran, Machine learning for high solid anaerobic digestion : Performance prediction and optimization, Bioresour. Technol. 400 (2024) 130665. https://doi.org/10.1016/j.biortech.2024.130665.

[67] Y. Ge, J. Tao, Z. Wang, C. Chen, L. Mu, H. Ruan, Y. Rodríguez Yon, H. Su, B. Yan, G. Chen, Modification of anaerobic digestion model No.1 with Machine learning models towards applicable and accurate simulation of biomass anaerobic digestion, Chem. Eng. J. 454 (2023) 140369. https://doi.org/10.1016/j.cej.2022.140369.

[68] O. Yildirim, B. Ozkaya, Prediction of biogas production of industrial scale anaerobic digestion plant by machine learning algorithms, Chemosphere 335 (2023) 138976. https://doi.org/10.1016/j.chemosphere.2023.138976.

[69] G. Baek, C. Lee, J. Yoon, Machine learning approach for predicting anaerobic digestion performance and stability in direct interspecies electron transfer-stimulated environments, Biochem. Eng. J. 193 (2023) 108840. https://doi.org/10.1016/j.bej.2023.108840.

[70] S. Il Yu, H. Jeong, J. Shin, S.G. Shin, A. Abbas, D. Yun, H. Bae, K.H. Cho, Simulation models of microbial community, pH, and volatile fatty acids of anaerobic digestion developed by machine learning, J. Water Process Eng. 60 (2024) 105225. https://doi.org/10.1016/j.jwpe.2024.105225.

[71] W.S. Kim, W.S. Lee, Y.J. Kim, A Review of the Applications of the Internet of Things (IoT) for Agricultural Automation, J. Biosyst. Eng. 45 (2020) 385–400. https://doi.org/10.1007/s42853-020-00078-3.

LIST OF PUBLICATIONS

My paper is currently under review in the **Renewable Energy Journal,** its impact factor is **9.0** and the date of submission was **15-11-2024**