Development of Artificial Intelligence Systems for Anaerobic Digestion Operations.

Offie, Ikechukwu Chukwudi (2023) Development of Artificial Intelligence Systems for Anaerobic Digestion Operations. Doctoral thesis, University of West London.

10.36828/xvqy2136

Development of Artificial Intelligence Systems for Anaerobic Digestion Operations

Ikechukwu Chukwudi Offie

A thesis submitted in partial fulfilment of the requirements of the University of West London for the degree of Doctor of Philosophy

Supervisory Team

Prof Kourosh Behzadian

Professor of Smart Infrastructure, University of West London

Prof Joe Rizzuto

Professor of Civil Engineering, University of West London

Prof Luiza Campos

Professor of Environmental Engineering, University College London

September 2023

ACKNOWLEDGMENT

This thesis completes a long process that began a few years ago. Along the way I walked with many people who were inspirational and incredibly supportive. I lack sufficient words to express my gratitude to everyone who helped me get to where I am and make my dream come true. However, I will try to thank those who played an important role in this journey.

First my gratitude goes to my principal supervisor, Prof Kourosh Behzadian for believing in me and constantly making efforts to encourage me especially in developing good research writing skills. These skills have really helped to boost my confidence in research writing.

A special gratitude goes to my fellow PhD colleague Dr Farzad Piadeh for taking time to put me through machine learning using MATLAB which made my research work relatively easy.

I would also like to thank Rokiah Yaman, Diego Vega and Dr Mark Walker who provided me with the data used to develop the models. I really do appreciate their support towards making my dreams come true.

Words are inadequate to explain my gratitude to my parents, Engr Okey, and Lady Kene for their financial support towards achieving my dreams. Leaving them to chase this future was perhaps the most difficult choice of my life, but they have supported me in every choice and thanks to them I am here now.

To my siblings, Uchenna, Ebele, and Daniel, thank them for always giving me a listening ear and constantly giving me words of encouragement during this journey.

Lastly, I would like to thank my wife, my best friend and companion Chidinma, for her patience, support, and encouragement. She has always believed in me with strong determination, celebrating my victories and regarding any failure as an accident. I could not have done this without her support and words of encouragement.

Abstract

This study explores two novel approaches for improving the performance of a micro anaerobic digestion system in generating maximum biogas. The micro anaerobic digestion system was a wet system situated in Camley-Central London. It operated continuously for 310 days under mesophilic conditions. The novel approaches include a new artificial intelligence-based model framework and an ensemble-based model framework. Both frameworks were developed using historic data obtained from the micro- anaerobic digestion system. The historic data include feed, catering, oats, liner, water, and biogas. The new artificial intelligence-based model framework entails developing a Recurrent Neural Network model for predicting biogas generated from the micro-anaerobic digestion system. The ensemble-based model framework entails combining different weak learning data mining models to improve the prediction accuracy of biogas generated. These weak learning data mining models include Support Vector Machines, K-Nearest Neighbour, Decision Tree, Gaussian Process Regression, Discriminant Analysis and Naïve Bayes. Both models were optimised after being trained to predict biogas using shuffled frog leaping algorithm to obtain the maximum biogas volume. The results showed great potential for the developed new artificial intelligence-based model in improving the performance of the micro anaerobic system in yielding optimal biogas by 43%. The results also showed that the average biogas produced could increase from 3.26 to 4.34 m^3/day. The developed ensemble model demonstrated 91% biogas prediction accuracy from the micro- anaerobic digestion system. The results of the weekly operation pattern led to 78% increase in biogas generation during the testing period. It also contributed to a 71% reduction in total required feeding days and 30% reduction in required pre-feeding days. The novel approaches demonstrated promising potentials in improving the performance of the micro-anaerobic digestion system to obtain maximum biogas with minimum energy and low operational costs making it a more viable option for managing organic wastes.

About the Author

Ikechukwu Chukwudi Offie acquired a Bachelor of Science degree in Civil Engineering at Eastern Mediterranean University Famagusta-North Cyprus in July 2015. He also acquired a Master of Science degree in Environmental Engineering at Cranfield University England in September 2016 and has been pursuing a PhD degree in Civil Engineering with the School of Computing and Engineering, University of West London, UK since September 2020. His research is centred on developing an artificial intelligence-based framework to monitor and improve the efficiency of anaerobic digestion operations. This was achieved using artificial intelligence (AI) based systems. His research also focusses on the real-time operation of anaerobic digestion using an ensemble model developed based on a combination of different weak learning data mining models. This was to improve anaerobic digestion operations in generating maximum biogas. Mr. Offie currently has three high impact journal paper publications, two conference papers.

Research Paper Publications

1. **Ikechukwu Offie**, Farzad Piadeh, Kourosh Behzadian, Luiza C. Campos, Rokiah Yaman (2023) **Development of an artificial intelligence-based framework for biogas generation from a micro anaerobic digestion plant.** *Journal of Waste Management*, 158. 66-75
2. Farzad Piadeh, **Ikechukwu Offie**, Kourosh Behzadian, Joseph Rizzuto, Angela Bywater, Mark Walker, Jose-Rodrigo Cordoba-Pachon (2023). **Assessment of the anaerobic digestion technology in achieving sustainable development goal: A review**. *Journal of Environmental Management*
3. Farzad Piadeh **Ikechukwu Offie**, Kourosh Behzadian, Luiza Campos, Angela Bywater (2023). **Real-Time Operation of Anaerobic Digestion Systems.** *Journal of Bioresource Technology* **(2023)**

Conference Papers

1. **Ikechukwu Offie**, Farzad Piadeh, Kourosh Behzadian, Luiza C. Campos, Rokiah Yaman (2022). **Real-Time monitoring of decentralized anaerobic digestion using Artificial Intelligence-based framework. *International Conference on Resources and Sustainability* (2022)**
2. **Ikechukwu Offie**, Farzad Piadeh, Kourosh Behzadian, Luiza C. Campos, Joseph Rizzuto (2023) **Real-Time Operation of Anaerobic Digestion Systems using an Ensemble Based Model. *Annual Doctoral Conference, University of West London (2023)***

Table of Contents

List of figures

List of tables

List of Abbreviations

AD = Anaerobic Digestion

ADM1 = Anaerobic Digestion No.1

AI = Artificial Intelligence

ALK = Alkalinity

ANN = Artificial Neural Network

BOD = Biological Oxygen Demand

BP = Back Propagation

COD = Chemical Oxygen Demand

C/N = Carbon to Nitrogen Ratio

CNN = Convolutional Neural Network

C-SVM = Conventional Support Vector Machine

DA = Discriminant Analysis

DT = Decision Trees

ELM = Extreme Learning Machine

FFNN = Feedforward Neural Network

FL = Fuzzy Logic

GA = Genetic Algorithm

GHG = Greenhouse gas

GLMNET = Generalized Linear Model Network

GPR = Gaussian Process Regression

GP = Genetic Programming

GBM = Gradient Boosting Machine

HRT = Hydraulic Retention Time

KNN = K- Nearest Neighbour

KPIs = Key Performance Indicators

KT = Kriging Technique

LS-SVM: Least Square-Support Vector Machine,

LMBP = Levenberg Marquardt Back Propagation

LSTM = Long Short-term Memory

LR = Linear Regression

ML = Machine Learning

MV = Machine Vision

MLP = Multi- Layer Perceptron

MPR = Methane Production Rate

MSW = Municipal Solid Wastes

MY = Methane Yield

NARX = Non-Linear Autoregressive Exogenous Input

NB = Naïve Bayes

NLP = Natural Language Processing

NN = Neural Network

NNSE = Normalized Non-Sutcliffe Efficiency

OLR = Organic Loading Rate

ORP = Oxidation Reduction Potential

OWM = Organic Waste Management

PSO = Particle Swarm Optimization

RBF = Radial Basis Function

RF = Random Forest

RMSE = Root Mean Square Error

RNN = Recurrent Neural Network

RNN-SMA = Recurrent Neural Network- Slime Mold Algorithm

SDGs = Sustainable Development Goals

SFLA = Shuffled Frog Learning Algorithm

SMA = Slime Mold Algorithm

SMY = Specific Methane Yield

SRT = Solid Retention Time

SSA = Sequential Sensitivity Analysis

SVM = Support Vector Machine

TAN = Total Ammonium Nitrogen

TN= Total Nitrogen

TP = Total Phosphorus

TS = Total Solid

VFA = volatile fatty acid

VS = Volatile Solid

XBOOST = Extreme Gradient Boosting

WLDM = Weak Learning Data Mining

ZVI=Zero Valent Iron

1 Chapter 1. Introduction

The world has continuously experienced rapid industrialised activities and economic development over the years (Usmani *et al.,*2020). These activities which have taken place over the years have led to exponential population growth and subsequent urbanization. It has also resulted in a tremendous increase in the amount of municipal solid wastes (MSW) generated annually. Recent studies by Kaza *et al* (2018) have shown that the world generates 2.01 billion tons of MSW annually with projections that it will increase to 2.2 billion tons by the year 2025 and to 3.88 billion tons by the year 2050.

The generation of MSW poses a multitude of challenges ranging from environmental problems such as ocean contamination, transmission of diseases, clogging of drains which causes flooding to social threats such as harm to human health and depletion of natural resources (Kumar *et al.,* 2021). It also causes economic losses as a study carried out by Wahba et al (2019) on Southeast Asia revealed that the economic cost of uncollected household waste that is burned, dumped, or discharged to waterways, at $375 per ton. The effect of this can lead to a decrease in both the value of dumping area and tourism (Saadatlu *et al*., 2023). These multitude of challenges occur mainly due to the poor management of these wastes, The effect of this has led to the emission of greenhouse gases into the atmosphere as studies by Kaza et al (2018) revealed that 1.6 billion tonnes of carbon dioxide (CO_2) equivalent greenhouse gases were emitted into the atmosphere from the volume of solid wastes generated in 2016, This represented about 5 percent of global emissions. Kaza *et al* (2018) also revealed that solid waste–related emissions are expected to increase to 2.6 billion tons of CO_2-equivalent per year by 2050 if no improvements are made in the sector.

Consequently, organic waste constitutes a significant fraction of the total MSW generated annually. It forms about 32% of the total MSW generated in high income countries, 53% in middle income countries and 57% in low-income countries. All regions generate about 50% or more organic waste on average except for Europe, Central Asia and North America where higher fractions of dry waste are generated (Kaza *et al*.,2018). This fraction of MSW has been recognised as a critical sustainability challenge, a fact underscored by its inclusion in the United Nations Sustainable Development Goals (SDGs) (Soni *et al.,* 2022). It is poorly managed as only 2% of the total amount of organic wastes generated are currently being treated and recycled (WBA.,2021). Following the poor management of organic waste which has contributed to a multitude of changes in the environment, society and economy, different nations and governments globally have been compelled to invest more financial and material resources in the remediation of organic waste in recent years (Wainaina *et al*., 2020).

Efforts are currently being made to revolutionize the waste management industry towards achieving sustainability and profitability (Abdallah *et al.*, 2020). This has led to the application of waste management technologies like anaerobic digestion (AD), composting, incineration, and landfill amongst others (Fazzo *et al*.,2020). The AD technology has been identified as one of the most effective techniques for the biological treatment of biomass sludges such as sewage sludge, food waste, agricultural waste etc. (Dos Santos *et al.,*2020; Rico *et al*;2020). The effectiveness of AD can be attributed mainly due to its ability to convert organic waste efficiently into valuable resources, thereby contributing to the growth of the economy while preventing the emission of GHGs into the atmosphere. Also, the ability of AD to prevent environmental pollution gives it an edge over other OWM techniques like landfill, incineration and composting which have been revealed to contribute to environmental pollution (Wainaina *et al.*,2020). The multi-faceted nature of AD has rendered it a highly ranked technique in the waste management industry and an excellent tool for the realization of circular economy (WBA., 2021). Despite the multi-faceted nature of the AD technology, its process is slow as it requires a long hydraulic retention time usually within the range of less than 5 days to 40 days depending on the type of digester (Obaideen *et al.,*2022;

Uddin & Wright 2022). The long hydraulic retention time of AD tends to raise the digester volume and cost. (Khalid *et al*., 2011). Also, the AD technology has been identified to have a low loading rate and slow recovery rate (Tang.,2003). These limitations hinder the wide application and adoption of the AD technology to full potential as they have been observed to directly affect the performance of AD in the production of biogas (Ankun Xu *et al*., 2021). To this effect, several mathematical models (theoretical, analytical, and statistical) such as AQUASIM, GRANIT BIOGAS, ANESSA and ADM1 have been developed with the aim of estimating and optimising the performance of AD (i.e., projection of biogas production and organic fertilizers) but their application has shown to have some limitations due to the complexity of their development, data demanding, and challenges associated with model calibration (Cruz *et al.* 2022). Also, the reliability of these models within the operation phase of AD plants has been observed to be more challenging as the operation conditions of AD processes can be highly variable and rapid changes in the control parameters are inevitable as it depends solely on waste composition (Cheela *et al*., 2021) .Based on this, the above-mentioned mathematical models are unable to give proper estimations of the model performance. Sequel to the shortcomings of mathematical models, data driven models such as artificial intelligence (AI) can be introduced as a good surrogate for process-based modelling as they are independent of complex physico-chemical processes. Various AI methods have been employed in AD systems for several purposes such as fault detection, process prediction, optimisation, and control of biological systems such as AD (Cruz *et al*.,2022; Wang *et al*.,2022). This is to ensure safety and improve the stability of the AD system (Kazemi *et al*.,2021). Several research works have studied the application of different AI methods in AD processes for modelling the relevant non-linear and complex relationships focusing on the optimization of particle size of organic matters, organic loading rate (OLR), ratio of carbon to nitrogen (C/N), pH and temperature, and retention time (Zhang *et al.,* 2019). However, to the best of author's knowledge, few research works have either presented an AI-based framework or an ensemble-based model for developing real-time operation strategies to improve the AD performance in producing biogas from the food waste generated in an urban area. Although the development of smart and decision-making frameworks in waste management have recently attracted more attention by researchers (Shahsavar *et al*. 2021; Shahsavar *et al*. 2022), none of the previous works either developed a framework for the operation of AD systems based on either RNN or a time series ensemble model for the real-time operation of AD. Also, no proper investigations were carried out on the effect of different waste compositions and the water added to the anaerobic digester on biogas yield. In addition, those previously developed models mainly used simple ML or ANN whereas the performance of the AD procedure may fit in better with simulation of time-series models that rely on earlier timesteps. This is particularly important because AD systems are operated continuously and are highly dependent on sequential and continuous input waste load (Yang *et al.,*2022; Chozhavendhan *et al*.,2023). This type of modelling can be envisaged through the application of a recurrent neural network (RNN) model for monitoring the performance and stability of the AD processes thus distinguishing it from previously developed AI models (Offie et al.,2022; Offie *et al.,*2023). Furthermore, the application of RNN model in AD has a shorter execution time, it does not require the multi-disciplinary knowledge related to bio-kinetics, microbiome, heat/mass transfer and avoidance of model re-calibration if trained based on extensive datasets compared to AQUASIM, ANESSA and ADM1 models. Hence, this study is aimed at developing an artificial intelligence (AI)-based framework for the optimal operation performance of an AD plant located in a residential area based on recurrent neural network (RNN) and optimisation techniques. It is also aimed at determining the maximum volume of biogas that can be generated from an AD plant using AI-based models. Moreso, a time series ensemble model will be developed using different weak learning data mining (WLDM) models to improve the biogas prediction accuracy of the AD plant. The developed ensemble model will then be optimised using an optimal algorithm, and the best input pattern which can yield maximum volume of biogas from the AD plant on a weekly basis will be investigated. This study will be achieved using data collected from a micro-AD plant in Camley Central London as the pilot study.

This thesis is organized as follows:

Chapter 1 presents an overview of the study. A background is provided, the problem/knowledge gap is stated and a summary of the current state of the research is presented. The aims and objectives of the study as well as the research questions and significance of the study are also stated.

Chapter 2 provides a detailed insight into the theoretical background of the subject. A definition of anaerobic digestion technology is given followed by a brief history of anaerobic digestion technology, its applications globally, the different types of AD systems as well as the challenges associated with the AD process. A critical evaluation of AD challenges will also be presented in this chapter. Moreso, a review of the various AI applications in anaerobic digestion systems for various purposes as well as the challenges and knowledge gaps will also be discussed in this chapter.

Chapter 3 describes the methodology adopted to achieve the aims and objectives of this research study. This chapter centres mainly on the series of steps taken to achieve the aims and objectives of this study.

Chapter 4 presents the results obtained from the proposed methodology outlined in this study which will be extensively discussed.

Chapter 5 presents the summary of research findings, the key contributions to knowledge and the relevance to the discipline. This chapter also presents the general recommendations for future research work based on the research findings which emerged from the analysis of the results obtained.

1.1 Research Aims

The research aims of this study are.

To develop an RNN model based on AI for improving the performance of the AD system in producing maximum volume of biogas. Secondly, to develop a time series ensemble model using different weak learning data mining (WLDM) models for the real-time operation of the AD system.

1.2 Research Objectives

The research aims of this study will be achieved through the following objectives.

1. To investigate the effectiveness of the developed RNN model in the accurate prediction of biogas produced from the AD system.
2. To determine the maximum volume of biogas that can be generated from the AD system using the developed RNN-SFLA model.
3. To determine the effectiveness of the developed time-series ensemble model for the real-time operation of the AD system.
4. To determine the optimal weekly pattern capable of yielding maximum volume of biogas from the AD system using the developed AI-based models (RNN and Time-Series Ensemble Model).

1.3 Research Questions

To address the objectives of this study, the following research questions will be addressed:

1. What impact does the developed RNN model have both on the AD system and other AD systems?
2. What is the significance of the maximum volume of biogas generated from the AD plant using the developed RNN-SFLA model on the AD system?

3. What is the significance of the developed time-series ensemble model on the AD system?
4. What impact does the determined optimal weekly pattern have on the overall operation of the AD system as well as other AD systems?

1.4 Scope of Study

Organic waste is a major source of concern in the world today. It has raised a lot of public health concerns over the years following the numerous environmental impacts associated with it. The effect of this has led to the introduction of AD technology which has been considered as one of the most effective organic waste management techniques. However, AD has its limitations which has hindered its application for full potential. This has necessitated the need for the introduction of AI and ensemble-based models into AD with the aim of improving its performance. Hence, this research study tends to address the limitations of AD using a new AI-based model and an ensemble-based model to improve its effectiveness in biogas production.

Table 1.1: Types of Anaerobic Digesters, HRTs, Temperature Range and TS contents

Type of Anaerobic Digester	Hydraulic Retention Time (HRT) (days)	Temperature Range	Total Solid Content (%)
Covered Lagoon	30-40	Psychrophilic	0.5-2
Complete Mix	0-25	Mesophilic/thermophilic	3-10
Plug Flow	10-25	Mesophilic/thermophilic	10-15
Fixed Film	Below 5	Mesophilic/thermophilic	1-5

Source: Uddin & Wright (2022)

Table 1.2: Linkage between the Research Objectives and Research Questions

Research Objectives	Research Questions	Cross-References
To investigate the effectiveness of the developed RNN model in the accurate prediction of biogas from the AD system	What impact does the developed RNN model have both on the AD system and on other AD systems?	Chapters 3 and 4
To determine the maximum volume of biogas that can be generated using the RNN-SFLA model.	What is the significance of the maximum volume of biogas generated from the AD plant using RNN-SFLA on other AD systems?	Chapter 4
To determine the effectiveness of the time-series ensemble model for the real-time operation of the AD system	What is the wider significance of the developed time-series ensemble model for the AD system?	Chapter 3 and 4
To determine the optimal weekly pattern capable of yielding maximum volume of biogas from the AD system using the developed AI-based models.	What impact does the optimal weekly patterns have on the overall operation of the AD system as well as other AD systems?	Chapter 4

2 Chapter 2. Literature Review

This chapter focuses mainly on the existing literature on the application of AI- and ensemble-based models in AD systems with the aim of improving their performance in the generation of maximum volume of biogas. To a large extent, an overview of anaerobic digestion technology is given laying emphasis on the history of the anaerobic digestion technology, types of anaerobic digestion systems, different designs, processes, parameters/factors influencing the AD process, its applications in the treatment of organic waste at different levels in different countries across the globe, its linkage to the SDGs and the challenges associated with anaerobic digestion technology and the challenges associated with the AD process will be presented in this chapter. A critical evaluation of AD challenges especially with regards to the management of the end products as well as the current stringent legislation on the reuse of the end-products will also be carried out in this chapter.

Lastly, a review of the concept of artificial intelligence, its brief history and applications in AD systems will be presented in this chapter. The challenges and knowledge gaps associated with the previous application of AI in AD systems will also be stated under this chapter.

2.1 Anaerobic Digestion (AD) Technology

Anaerobic Digestion is a process in which microbes break down or digest complex organic waste into another form in the absence of oxygen (Obaideen *et al*.,2022). The diverse microbial population degrades organic waste resulting in the production of biogas and other energy-rich organic compounds such as liquid and solid fertilizer highly useful to mankind in numerous ways as end products (Azeem *et al*.,2011), as observed in Figure 2.1. It has been identified as a green technology which has been applied for more than 100 years to stabilize organic waste including sewage sludge, food waste and livestock wastes to produce methane gas/biogas as a by-product (Cruz *et al.,*2022). The history of AD can be traced back to the 10th Century BC and 16th Century in Assyria and Persia respectively where biogas was used to heat bath water (Auer *et al.,*2017). In the 17th Century, Jan Baptita Van Helmont discovered that flammable gases could evolve from decaying organic matter. In 1778, the physicist Alessandro Volta scientifically identified that gas as methane (Gijzen., 2002). Later in 1808, Sir Humprey Davy discovered that methane was present in the gases produc^ed during the anaerobic digestion of cattle manure and in 1859, the first digestion plant was built at a leper colony in Bombay, India.

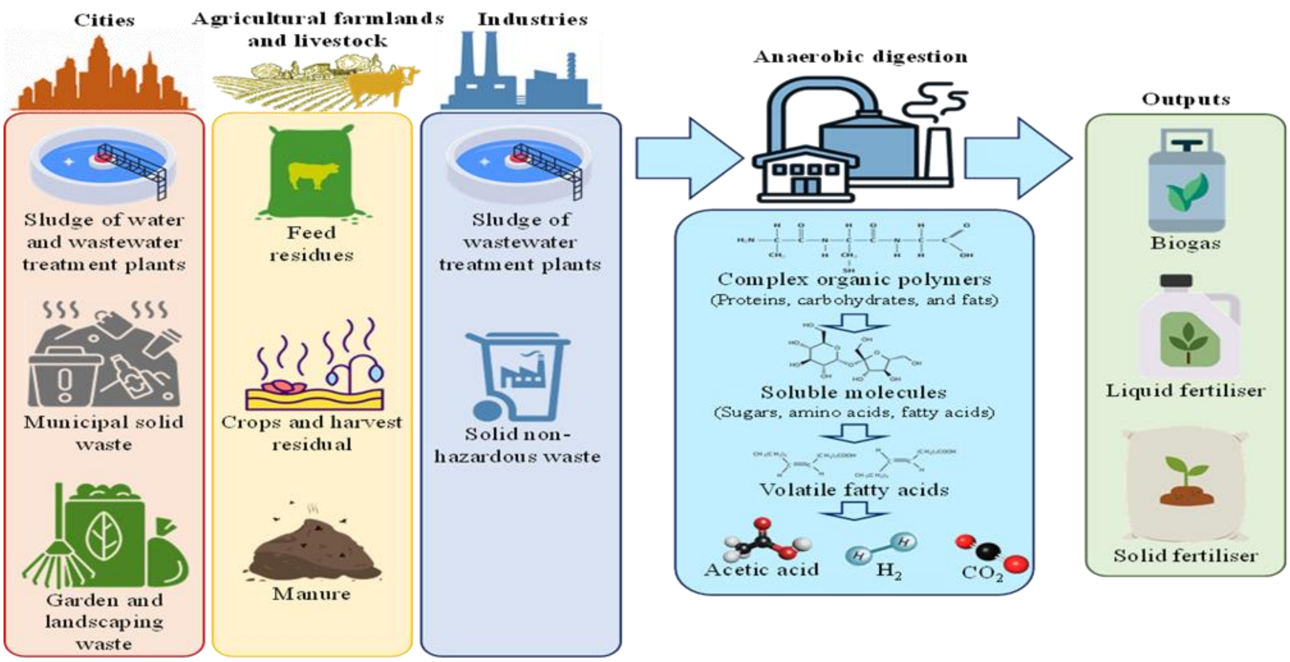


Figure 2.1: Anaerobic digestion technology and its applications

In 1895, anaerobic digestion got to England when biogas was recovered from a carefully designed sewage treatment facility and used to fuel streetlamps in Exeter. This then resulted in the development of microbiology as a science, which led to research by Buswell and others in the 1930s to identify anaerobic bacteria and the conditions that promote the production of methane. The AD technology rose to prominence following the shortage of fuel during World War II (1939-1945) where it was used to produce fuel from biogas. However, interest in AD dropped after the World War II until in the 1970s where it sparked interest as a result of the global energy crisis.

The AD technology has been identified as one of the most attractive methods for the treatment of both organic liquid effluents and organic solid wastes (Dhussa *et al.*,2014). It has continuously received attention globally over the years having been successfully implemented both at the industrial and domestic level playing a vital role in our world (Karagiannidis & Perkoulidis.,2019). It has proven to be highly useful in the areas of biowaste management, food production, energy production and pollution prevention.

2.1.1 Types of AD Systems

AD technology is made up of different types which exist based on the moisture content of the feedstock, feeding frequency, mixing type and temperature (Uddin & Wright 2022). This is illustrated in Figure 2.2. The moisture content of the feedstock is divided into two categories namely wet AD and dry AD. In the wet AD system, the moisture content of the feedstock is more than 85%. Feedstocks are stirred mechanically to prevent solid precipitation. Generally, substrates are continuously fed to the anaerobic digester and removed after a specific HRT has been attained. Feedstocks containing high moisture content such as sewage sludge and animal manure adopt wet AD process due to the high energy demand required to reduce their moisture content. Dry AD is used for feedstock having a higher solid content above 15% (Uddin & Wright 2022). The feedstocks are stacked in a sealed tank with hot water or slurry spread over the feedstock to provide a specific digestion temperature. Substrates such as solid animal manure, biosolids from municipal solid waste (MSW), food waste, yard trimmings, and energy crops are suitable for the dry AD process.

The feeding frequency is subdivided into two categories namely, A batch digester and the continuous digester. In the batch digester, feedstocks are added at the beginning of the process and kept covered for a specific period. The digester is then emptied prior to the addition of the next batch of feedstocks. Operation and maintenance of a batch digester is simple, but the production of biogas is periodic. However, in the continuous anaerobic digester, feed-stocks are continuously added with biogas and digestates being removed at a similar rate. Continuous digesters constantly produce biogas with minimum digester downtime.

In the case of mixing, the feedstock can either be completely mixed or not mixed at all. Mixing of feedstock can be carried out in different ways, such as mechanical agitation, biogas recirculation, recirculation of digesting content using either a pump or nozzle. The mixing process requires a complex design of the digester (Uddin & Wright.,2022). The operating costs of the mixing category are usually higher compared to the non-mixing category.

Lastly, the temperature of AD is categorized into three different types namely psychrophilic, mesophilic, and thermophilic. The psychrophilic temperature occurs at temperatures less than 20 degrees Celsius. Mesophilic temperature occurs at temperatures within the range of 30-45 degrees Celsius. The thermophilic temperature operates at temperatures within the range of 55-60 degrees Celsius (Uddin & Wright.,2022). However, previous research studies carried out on anaerobic digestion technology have shown that anaerobic digestion operations are mostly carried out at mesophilic temperatures (El-Mashad *et al*., 2003; Khalid *et al.*,2011). The operation in the mesophilic range is more stable thereby requiring a smaller expense of energy (Fernandez *et al.,* 2008; Ward *et al.,* 2008). Another study by Castillo *et al*. (2006) revealed that the best operational temperature was 35˚C with an 18-day digestion period. The study by Castillo *et al* in 2006 also revealed that a little fluctuation in temperature from 35 ˚C to 30˚C caused a reduction in the rate of biogas production (Chae *et al*., 2008). Generally, a temperature range between 35– 37˚C is considered suitable to produce methane and a change from mesophilic to thermophilic temperatures can cause a sharp decrease in biogas production until the necessary populations have increased in number (Khalid *et al.,*2011).

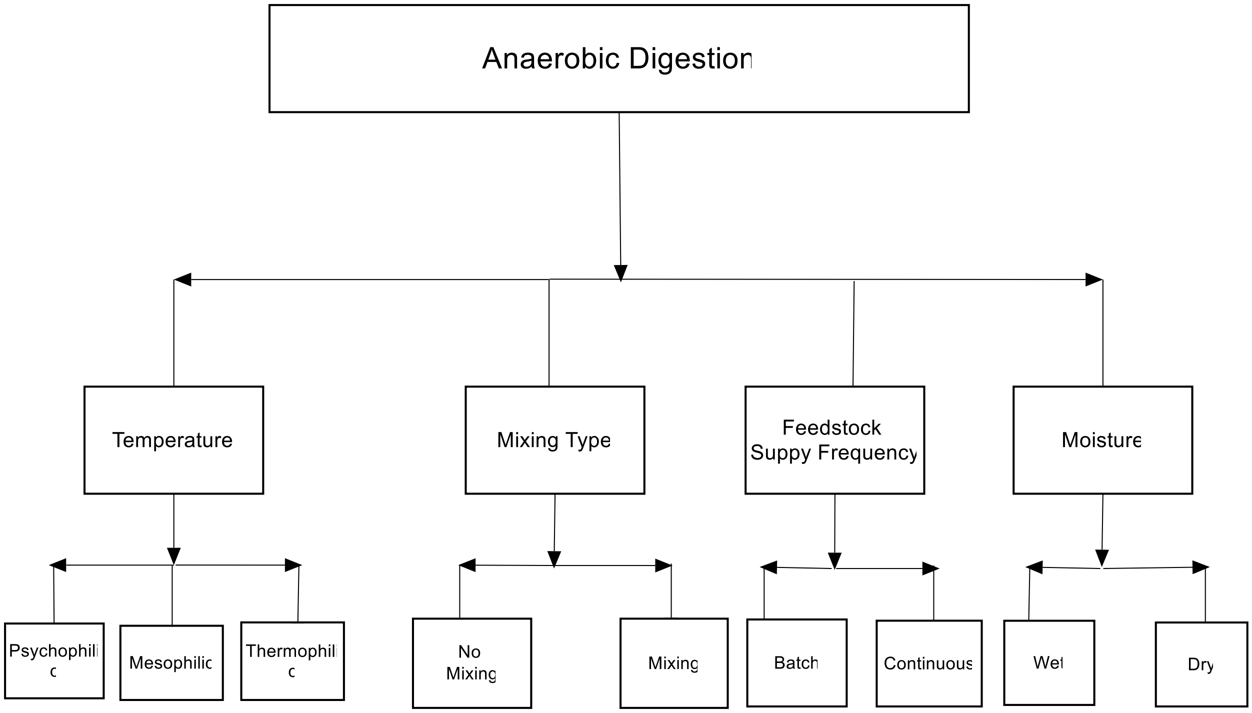


Figure 2.2: Structure of the types of AD Systems

2.1.2 Types of AD Designs

The AD system has different standard digester designs. These standard digester designs include, covered lagoon, complete mix, plug flow and fixed film.

In covered lagoons, feedstocks are stored in an underground lagoon covered with a gas-tight flexible cover. The lagoon serves simultaneously as storage and reactor. Covered lagoon digesters are best suitable in warmer regions where the ambient temperature is sufficient to provide the required digestion temperature. Feedstocks having low solid content within range 0.5–2% are optimal for this type of digester due to the easy and inexpensive handling of larger volumes. The typical HRT is 30–45 days. Often, screening larger solid particles from the feedstock is necessary to prevent a crust from forming on the lagoon surface to lower the biogas production efficiency (Uddin & Wright 2022).

A complete mix digester Is an above-ground tank made of insulated concrete or steel. A rigid or flexible cover is used to hold the produced biogas and later collect via gas collecting pipes. Heat exchangers maintain the digestion temperature, and generally, a mechanical mixing system is attached to ensure complete mixing of the feedstock. Complete mix digesters are capable of handling non-homogeneous feedstock with higher solid content (3–10%) feedstock. Like covered lagoons, they are suitable for any ambient conditions. The HRT for complete mix digester is lower than for a covered lagoon ranging from 10 to 25 days.

Plug flow digesters operate similarly to the complete mix digester, except for the feedstock having no mechanical mixing. The plug flow digester is a horizontal, cylindrical shape reactor where feedstock enters from one end and the digestate exits from the other end. The incoming feedstock pushes out an equal amount of substrate while digestion occurs along the way. Plug flow digesters are typically in-ground and covered with a flexible cover. The feedstock solid content needs to be high (<10–15%) to ensure the movement of fluid through the reactor (Uddin & Wright.,2022).

Fixed film digester design promotes microbial growth as a thin film on the surface, often called a biofilm. A column packed with supporting media such as a small plastic ring or wood chips is placed inside the digester. This type of digester is not suitable for all substrates as the packed column has a very narrow space for the substrate flow. The acceptable solid content for this type of digester is 1–2%; higher solids can clog the substrate flow through the digester media. A shorter HRT, typically 2–6 days, is the main characteristic of this type of digester, resulting in a smaller digester volume.

2.1.3 Processes of the AD system

The anaerobic digestion of organic material is a complex and multi-step process (Khalid *et al.,*2011). It involves several different degradation steps. The microorganisms that participate in the process may be specific for each degradation step and thus could have different environmental requirements (Gupta *et al.,*2022). These stages include enzymatic hydrolysis, acidogenesis/fermentation, acetogenesis and methanogenesis (Park *et al.*,2005; Charles *et al*.,2009) as shown in figure 2.3. A pre -treatment step is usually carried out to precede the actual biodegradation process. This is to adequately prepare substrates with different structures such as lignocelluloses (Sarsaiya *et al*., 2019a). It also helps to improve the quality of biogas generated as the quality of biogas generated from an AD plant is dependent on the quality of the substrate/waste input. The enzymatic hydrolysis stage is a chemical process where, complex organic molecules are converted into a simple substance like amino acid, long-chain carboxylic acid, and sugars (Rasapoor *et al*., 2020). Fermentation/Acidogenesis is a biological process in which bacteria are used to decompose the simple monomers to sugars and amino acids into different by-products like ammonia, hydrogen, organic acids, and carbon (Pramanik *et al*., 2019). Similarly, acetogenesis is the biological reaction, where volatile fatty acid (VFA) is converted into amide ion, hydrogen, and carbon dioxide. Methanogenesis is a biological process where methanogens are used to convert digested materials into methane and carbon dioxide which forms biogas (Arif *et al*., 2018).

Due to the AD process which involves a series of biochemical and physical processes, its efficiency and stability are influenced by various parameters depending on the type of the AD plant. These parameters include temperature, pH, moisture content, carbon to nitrogen ratio (C: N), organic loading rate (OLR), hydraulic retention time (HRT), Total Solids (TS) and Volatile Solids (VS) (Khalid et al., 2011; Uddin & Wright 2022). The temperature parameter has a significant effect on the microbial community, kinetics of the process and stability as well as the methane yield from the biogas produced). For instance, low temperature during the AD process has been observed to decrease the growth of microbes, the rate of substrate utilization as well as the production of biogas (Trzcinski & Stuckey, 2010). It could also result in an exhaustion of cell energy, a leakage of intracellular substances or complete lysis (Kashyap *et al.,* 2003). In addition, high temperatures cause a decrease biogas yield due to the production of volatile gases such as ammonia which suppresses methanogenic activities (Fezzani and Cheikh, 2010). Generally, anaerobic digestion operations are carried out at mesophilic temperatures (i.e., temperature between 35˚C-37˚C). This is because AD operation in the mesophilic range is more stable and requires a smaller energy expense compared to the thermophilic range (Fernandez *et al*., 2008). Also, temperature within this range (mesophilic) ensures suitable production of both methane and biogas from AD.

The pH parameter has a significant effect on the microbial degradation efficiency which influences biogas yield produced from AD operations (Jayaraj *et al.,*2014). Moisture content influences the performance of the degradation process by dissolving readily degradable organic matter (Khalid *et al*.,2011). The carbon to nitrogen ratio of the organic material plays a crucial role in achieving a balanced AD process. The carbon ratio is the energy source for microorganisms, while the nitrogen ratio is required for the microbial growth or metabolism. The OLR describes the input rate of the organic material per unit volume of the anaerobic digester. It is dependent on the concentration of the substrate organic matter. OLR is also a critical operational parameter influencing biogas production from AD as it represents the biological conversion capacity of the AD system (Sun *et al.,*2017). The HRT variable is another important variable/parameter influencing AD operations. It has a significant effect on the performance of the anaerobic digester in producing biogas efficiently (Ezekoye *et al.,*2011). However, its significance in the performance of the anaerobic digester in producing biogas efficiently is dependent on the temperature of the digester. Also, it plays a major role the efficiency of the AD system as literature has shown that the AD system is inefficient at lower OLR. It improves with an increase in OLR at an appropriate range. On the other hand, the increase in OLR beyond its appropriate range leads to a drastic decrease in biogas yield. This causes a corresponding failure of the AD system (Li *et al*.,2015). Total Solid (TS) of the feedstock plays a vital role in the performance of the AD plant. According to studies carried out by Yi *et al* (2014), change in the content of total solids will lead to a change of microbial morphology in systems which had a significant impact on the performance of the AD system. For instance, an increase in the total solid content led to a better performance of the AD system including reduction of volatile solids and methane yield. The Volatile Solid (VS) is a fraction of the TS mostly within the range of 89-92%. Like TS, it provides useful information about the yield of biogas expected from the AD process as well as the efficiency (Orhorhoro *et al*.,2017).

2.1.4 **Link between AD and SDGs**

The application of AD technology for biowaste management, food production, energy production and pollution prevention purposes cuts across the three pillars of sustainable development (i.e., environment, society, and economy). These three pillars of sustainable development form the basis of human existence (Obaideen *et al*.,2022). It can operate at sizes from that of a test tube to tanks of many thousands of cubic meters (WBA., 2021). As such, it is adaptable and has been revealed to play vital roles in contributing towards the achievement of 15 out of the 17 sustainable development goals (SDGs) both directly and indirectly in the remotest parts of the global south to the organic wastes created by world cities such as New York (Obaideen *et al*.,2022). The various contributions of AD towards the achievement of the 15 out of the 17 SDGs are presented in Table 2.1.

An example is the application of AD for the purpose of providing clean and affordable energy in line with contributing towards achieving SDG 7. The application of AD for this purpose has been observed both in developing countries and developed countries across the globe. For instance, countries like China, India, Bangladesh, Nepal, Vietnam, Sri Lanka, Pakistan and Thailand have extensively applied on AD on small scale for household cooking, heating, and lighting (Eurostat & European Statistics.,2017; N. Scarlet.,2015). Reports by Fairfield in 2021 revealed that there were more than 6 million anaerobic digesters in China and India alone where most of the anaerobic digesters are small scale. It has also been extensively applied in developed countries within the EU where they have been used for advanced purposes and are currently the world leader in the production of biogas for electricity from more than 17,400 biogas plants. In addition, AD technology has also been used in the United States where there were more than 2100 biogas plants in 2017, of which 250 farm-based digestion plants using livestock manure (US, EPA.,2017). Also, about 1240 WWTPs operating anaerobic digesters producing biogas from 15,000 WWTPs were reported to exist in the US in 2017. In the case of Africa, biogas production from AD plants is still less developed compared to other regions globally however, biogas digesters have been installed in several African countries like Burundi, Botswana, Burkina Faso, Cote d'Ivoire, Ethiopia, Ghana, Guinea, Lesotho, Kenya, Namibia, Nigeria, Rwanda, Senegal, South Africa, Uganda, and Zimbabwe) through the implementation of national programs (Viancelli *et al*.,2019; Cheng S *et al.*,2014). These programs have been successfully implemented in Rwanda, Tanzania, Kenya, Uganda, Ethiopia, Cameroon, Benin, and Burkina Faso (Austin & Morris 2012). Furthermore, a Biogas Partnership Programme (ABPP), supported by the Ministry of Foreign Affairs of the Netherlands and Netherlands Development Organisation aimed at developing national biogas programs in five African countries (Ethiopia, Kenya, Tanzania, Uganda, and Burkina Faso) was established for building 100,000 domestic digestors to provide access to clean energy for a half million people by 2017. The program led to the installation of almost 60000 biogas plans in these five countries (16,419 in Kenya, 13,584 in Ethiopia, 13,037 in Tanzania, 6504 in Uganda and 7518 in Burkina Faso) as reported in 2017 (N. Scarlet *et al*.,2018). Several agricultural and domestic biogas plants have also been set up for rural households in Latin America. Further studies conducted by N. Scarlet *et al* (2018) revealed that the network for Biodigesters in Latin America and the Caribbean have promoted the development of small bio-digesters in Bolivia, Costa Rica, Ecuador, Mexico, Nicaragua, and Peru where Bolivia was reported to be leading with over 1000 domestic biogas plants installed. Large-scale biogas plants have been built to use effluents from palm oil mills and large farms in Colombia, Honduras, and Argentina (Kapoor & Vijay 2013). Brazil was reported to have 127 biogas plants using agricultural and industry residues, biowaste, sewage sludge, and landfill gas, which produced about 1.6 million Nm^3 /day, (584 billion m3 biogas/year) representing 3835 GWh of energy in 2015 (IEA.,2015; N. Scarlet *et al.*,2018). The installed biogas electricity capacity has increased significantly in the last years, reaching 196 MW in 2015 and 461 MW in 2016 (IRENA.,2023). The installed biogas electricity capacity was also observed to increase to 486 MW in 2022(IRENA.,2023).

Through the application of AD technology across these countries, it helps to prevent environmental pollution as the AD technology serves as an alternative to household energy sources such as wood, coal, and liquefied petroleum gas which have adverse effects on the environment through the burning of fossil fuels (a major contributor to global warming which exacerbates the variability of climate) (Hoppe and Sander, 2014).It has also helped to improve the quality of living standards through the provision of better air quality and better health as the production of biogas from the use of AD for domestic purposes has helped to replace the use of firewood which produces a lot of smoke and soot particles harmful to the human health (Gautam *et al*.,2009).

Another research study of the application of AD for 12 remote families in a project to replace firewood with biogas indicated that firewood usage decreased by 50-60%, and the cooking time was reduced by 1hr (Garfi *et al.,* 2016). The resultant effect of this led to a reduction in the burning of fossil fuels and deforestation which are major contributors to global warming, climate change and melting of the polar ice which have adverse effects on the human health. Moreover, the use of biogas for cooking has helped to reduce the burden of women in remote families charged with the responsibility of going to long distances to collect firewood used for cooking (WBA.,2018). This is in line with the achievement of SDGs 3 and 5 respectively.

Furthermore, the application of AD for the purpose of generating renewable energy has demonstrated to have positive impacts on the environment, society and economy, contributing towards the achievement of SDGs 4, 6,8,9,11,12, 13,14, 15 and 16. This observation has been made by different research works that have been carried out on the AD technology in

Bangladesh where showing the contributions of AD towards improving the quality of education in rural communities through increasing energy accessibility thereby contributing towards achieving SDG 4. This assessment of a study about cow dung used to generate biogas and its effect on the sustainable development of a district in Bangladesh showed that the biogas plant provided an efficient way of converting cow dung into useful energy and fertiliser (Shaibur *et al*., 2021). It also enhanced the cooking environment for biogas digesters, which decreased the time required to collect wood for cooking food, providing people ample time to attain education.

The application of AD for energy generation has contributed towards economic growth, establishment of small-scale industries in rural areas as well as sustainable cities and communities in different parts of the world. These are in line with achieving SDGs 8,9 and 11. These have been observed in China, India, United States, France, Italy, United Kingdom, and Germany where series of digesters have been constructed and used for treating different kinds of organic wastes (Akhiar *et al*., 2020). In addition, it has helped to enhance the efficiency of natural resource usage and improving the waste recycling process as more organic wastes have been diverted from landfills. This is in accordance with achieving SDG 12. The diversion of more wastes from landfills has also helped in contributing towards the mitigation of climate change, in accordance with achieving SDG 13. It has contributed towards minimising air pollution from the emission of GHGs such as methane and carbon dioxide.

Moreover, it has helped to promote environmental sanitation globally in line with contributing towards achieving SDGs 6, 14 and 15 through water and soil pollution prevention which occur from landfill leachate. The establishment of AD programmes in rural communities can also help to unite people as it brings them together for the achievement of a common goal (WBA, 2018). Through this means, peace and unity amongst the people are promoted as the establishment of AD technology for biogas and organic fertiliser production have several communal and societal benefits. This is in line with achieving SDG 16 (Winter.,2008). Furthermore, AD technology has been found useful in the area of food production through the promotion of sustainable agriculture in accordance with achieving SDG 2 (Obaideen *et al*.,2022). Though it is still emerging in developing countries around the world however, reports have shown that the application of AD in agriculture has significantly increased the carbon content of the soils (Lohani *et al*.,2021). It has also led to a better soil coverage, reduced leaching, and run-off due to self-sufficiency in fertilisers (Bong *et al.,* 2017). Reports have also shown that adding value to the waste through the application of AD has helped to change it from a burden on the government into an opportunity to produce biogas, bio-fertiliser and create new jobs especially in China and United States with the aim of achieving SDG 1 (Obaideen *et al.,* 2022).

The AD technology is a well-established technology for the treatment of both solid and liquid organic wastes (Dhussa *et al*.,2014). It is regarded as a more preferable waste management option than others in the organic waste management industry. This is because it has a low energy requirement and a low biomass for operation unlike the composting technology. Also, the ability of AD to prevent environmental pollution as it takes place in the absence of oxygen makes it a more preferable and sustainable option for managing organic wastes compared to other organic waste management techniques like landfill, incineration, and composting.

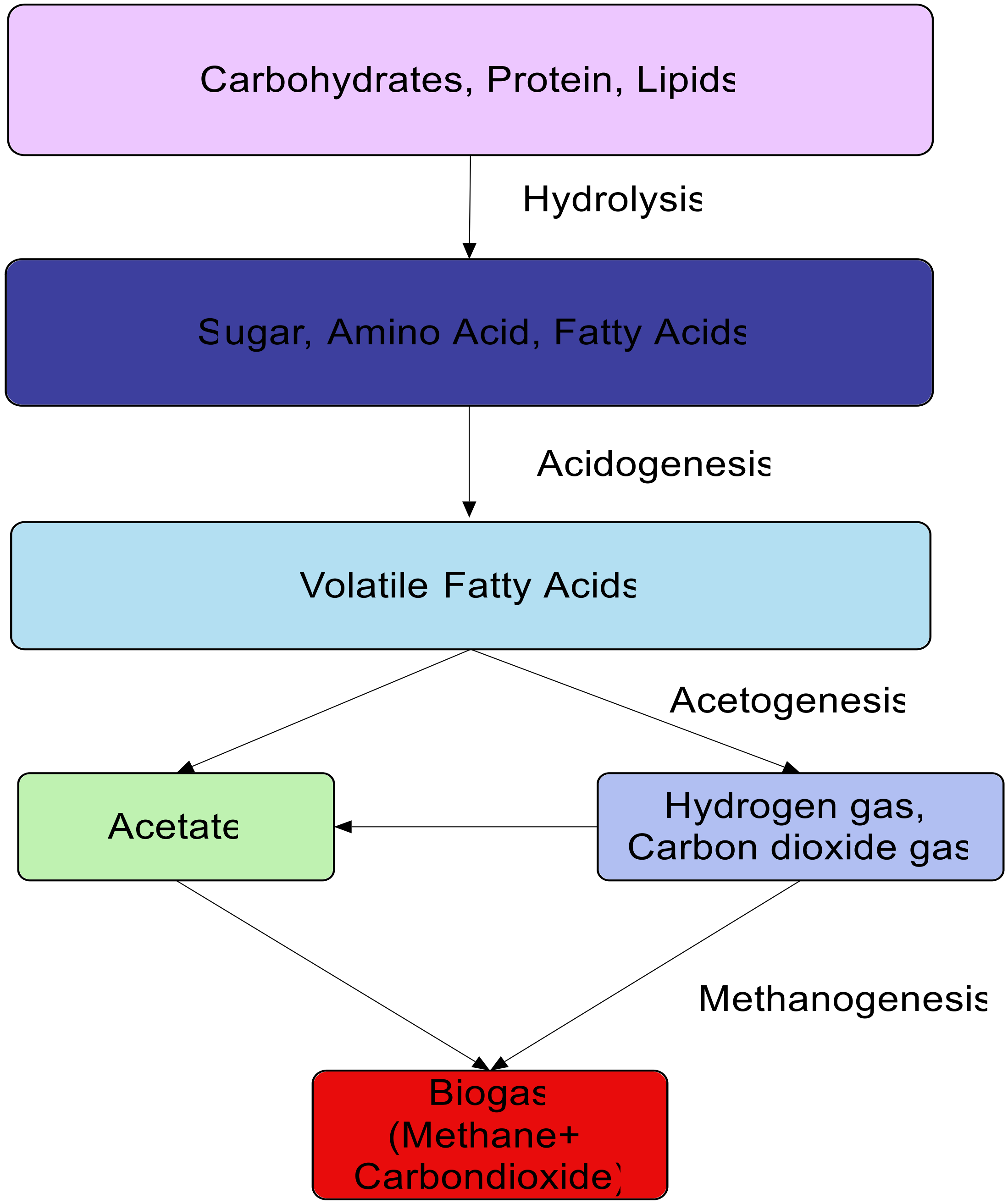


Figure 2.3: Processes in Anaerobic Digestion Technology

Table 2.1: Connection between the potential contribution of AD to the SDGs

	SDG	Objective	Contribution of AD to SDG	References
1.	**No Poverty**	To eradicate poverty in all its forms everywhere across the globe.	Assisting the smallholder by providing affordable fertilizer and eliminating the issues of the complexity of the fertilizer supply chain. Creation of jobs through the adoption of a new business	Herrmann., 2013
2.	**Zero Hunger**	End hunger, achieve food security and improved nutrition. To promote sustainable agriculture everywhere across the globe.	Restoring soils through the recycling of nutrients, organic matter, and carbon Increasing the yield of crop through the use of nutrient-rich digestate bio-fertilizer. Recirculating phosphorus, which is vital for the growth of plants but limited in supply.	Arthurson., 2009
3.	**Good-Health and Well-Being**	To ensure healthy lives and promote well-being for all ages.	Reduction of indoor air pollution through the substitution of solid biomass-based domestic fuels with biogas Diverting organic wastes from landfills by treating and recycling organic wastes will help to reduce odours and the spread of diseases.	(Ilo *et al.,* 2020; Zeng *et al*., 2020)
4.	**Quality Education**	Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.	Increasing energy accessibility in rural areas will improve the quality of education.	Obaideen *et al*., 2022
5.	**Gender Equality**	To achieve gender equality and empower all women and girls globally	Reducing the burden of collecting firewood in remote areas with the aim of improving the quality of women and children's lives.	Tamburini *et al.,* (2020)
6.	**Clean Water and Sanitation**	To ensure the availability and sustainable management of water and sanitation for all	Providing decentralized, local treatment of biosolids in remote and rural communities to reduce odours and the spread of disease. Stabilizing and recycling biosolids through anaerobic digestion to enable it to be applied back to land. Reducing the carbon loading of wastewater to reduce its impact on water bodies.	Adnan *et al.,* 2019
7.	**Affordable and Clean Energy**	Ensure access to affordable, reliable, sustainable, and modern energy for all	Reducing the dependence on fossil-fuel based energy sources by replacing with biogas. Capturing waste heat from co-generating units linked to biogas plants. Utilizing locally produced waste and crops to generate energy for rural and remote communities. Storing biogas to produce energy when required	Ullah *et al*., (2017)

8. Decent Work and Economic Growth	To foster sustained, inclusive, and sustainable economic growth, full and productive employment	Increasing the gross domestic product (GDP) by enhancing waste utilization. Reducing materials carbon footprint	(Ahmad *et al*., 2018 Adnan *et al*., 2019; Hansupalak *et al*., 2016; Okoro & Sun., 2019)
9. Industry, Innovation, and Infrastructure	To build resilient Infrastructure, To promote sustainable industrialization and foster Innovation	Improving the self-sufficiency and sustainability of industries by extracting the energy from their own effluents Collaboration between industries and agriculture for mutual benefit Generating short term construction employment and long-term equipment manufacturing and maintenance employment Encouraging the growth of micro-enterprises through the provision of reliable electricity	(Verhoog *et al.,* 2016) (Abdul Aziz *et al*., 2019)
11: Sustainable Cities and Communities	To make both cities and human settlements safe, resilient, and sustainable	Meeting the needs of people for basic services including energy and water. Reducing the adverse environmental impact of cities by investing in renewable energy, managing scarce resources, and improving waste and recycling systems.	(Yasar *et al.,* 2017), (Velivela *et al*., 2020) (Kelebe *et al*., 2017)
12: Responsible Consumption and Production	Ensuring sustainable consumption and production patterns	Enhancing the efficiency of natural resource usage. Reducing air and water pollution Improving the waste recycling process	Paolini *et al*., 2018, (Jeong *et al*.,2017)
13: Climate Action	To take urgent action towards the mitigation of climate change and its impacts	Reducing the emission of GHGs through the provision of a lower-emission energy source. Reducing methane emissions from the livestock industry. Reducing GHG emissions from landfills.	Lima *et al.,* 2018
14: Life Below Water	To conserve and sustainably utilize oceans, seas, and marine resources for sustainable development.	Reducing marine pollution by preventing land source pollution Improving the freshwater ecosystems through the enchantment of wastewater treatment.	Khaled *et al.*,2022
15: Life on Land	To protect, restore, and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, halt and reverse land degradation and halt biodiversity loss	Recirculating nutrients and organic matter in organic wastes through anaerobic digestion and returning them to the soil in the form of digestate biofertilizer. Substituting firewood with biogas as a domestic fuel helps in reducing deforestation. Improving land-use productivity and reducing land-use change.	(Studer *et al.,* 2017 ; Tamburini *et al.,* 2020)
16: Peace and Justice	To promote peaceful and inclusive societies for sustainable development, provide access to justice for all	Some research studies have shown that the increase in power availability is directly linked to peace. Hence the	(Winter., 2008)

Strong Institutions	and build effective, accountable, and inclusive institutions at all levels	establishment of AD in communities lacking adequate power supply helps to unite people in communities, bringing them together for the achievement of a common goal.

2.1.5 Challenges associated with the AD processes and how they can be addressed.

Despite the numerous potentials of AD technology, it still suffers from various challenges which have hindered its widespread application. Some of the significant challenges associated with the AD process include feedstock variability, low process efficiency, and process instability amongst others (Uddin & Wright.,2022). The instability of the AD process affects biogas production and sometimes leads to failure of the process. Following this, a wide range of techniques have been identified suitable for addressing these challenges. These techniques include thermal, chemical, and mechanical pretreatment methods.

Different thermal, chemical, and mechanical pretreatment methods have been identified to improve the hydrolysis or solubility of the digester's organic materials (Uddin & Wright.,2022). For instance, conventional heating of the substrate, also known as the thermal process, increases the solubility of organic materials in water. It also provides pathogen-free feed to avoid process inhibition. This technique is specifically useful for industrial-scale wastewater treatment. Also, microwave irradiation has being identified as a low-energy alternative. This technique uses focused direct heat to improve the degradability of complex polymers. In the case of lignin-rich substrate, the addition of acids or bases can improve solubility and enhance the production of biogas. Though this process is energy and cost-intensive, the addition of oxidants is useful when the waste substrate mainly consists of recalcitrant components such as lignin (Uddin & Wright.,2022).

Mechanical pretreatment methods such as grinding, shredding, milling, and screening are mostly used to improve the efficiency of the digestion process. This method mainly increases the molecule surface area and enhances bacterial activity during the digestion process. Another mechanical pretreatment method used in AD is high pressure homogenization (HPH). This method has been found to be useful in homogenizing the substrate where the substrate cell membranes are disrupted using high pressure induced shear ranging between 30-150MPa. Also, process inhibitions which occur due to the accumulation of harmful intermediate products and nutrients imbalance can be minimized through different techniques. However, the optimisation of OLR is a well-known approach used in reducing the accumulation of VFA. In addition, codigestion or adding other organic materials assists in the maintenance of nutrient balance. It also helps to avoid process inhibition. Codigestion has also been revealed to be applicable for ensuring optimum C/N ratio (Uddin & Wright.,2022).

The chemical pretreatment method involves the application of additives into the AD system. Additives are used in the digester to improve material conversion and biogas production. It supports the growth of microbes, adsorption of inhibitory products, nutrient supplementation, and enhancing buffering capacity (Arif *et al.,*2018). Various conductive materials such as sand, molecular sieve, zeolite, charcoal, etc., are used as additives to improve syntrophic activity while creating an enabling environment for microbial growth (Yang *et al*.,2016). These materials can also absorb inhibitory products like NH_3 and H_2S resulting in more efficient conversion. If any substrate lacks specific nutrients necessary for the digestion process, micro- and macro-nutrient supplements are added to the anaerobic digester (Uddin & Wright.,2022). The addition of additives stimulates biogas production while helping to maintain process stability (Romero-Guiza *et al.,*2016).

However, it is important to note that the suitable method for addressing these challenges associated with the AD process solely depends on the type of feedstock, digestion technology and targeted outcome. This is because not all types of techniques are applicable for every AD technology.

2.1.6 Critical Evaluation of AD Challenges with Specific Emphasis on End Product Management

The AD process is a biochemical process as it involves a series of biological and chemical processes as discussed on 2.1.1. These series of processes lead to the production of biogas and digestate. Biogas consists of methane (45%-75%) as a renewable energy source, carbon dioxide (25%-55%), and small amounts of hydrogen sulfide (H_2S) and hydrogen (H_2) (Panuccio *et al*., 2016). Digestate is the remaining slurry formed from the AD process. Depending on the composition of feedstock and the design of AD system, about 20–95% of the organic matter is broken down to produce these end products (Moller and Müller, 2012).

These two products of AD have several environmental, social, and economic benefits which have made AD technology a preferable option for the treatment of municipal organic wastes. However, despite the numerous benefits associated with the two end products of AD, they have been revealed to have negative environmental impacts which are harmful to human health in various ways. For instance, digestate has the potential to cause harm to both the growth of plant and the soil which can lead to problems for its sustainable disposal due to its chemical composition (Rigby and Smith, 2013), Also, the early application of digestate and its longer retention time in the soil without usage by crops has the potential to cause the loss of nutrients and their translocation towards deeper soil layers or the emission of NO_3 into groundwater (Formowitz and Fritz, 2010). Moreover, digestate pH values above 8 could lead to additional volatilization losses (Formowitz and Fritz, 2010).

In the case of biogas, one of the major challenges facing its application is contamination with different critical impurities such as sulfur compounds, siloxanes, halogens, VOCs, ammonia etc (Obaideen *et al.,*2022). These impurities are dependent on the feedstock used in biogas production. They have various negative impacts which even after their purification, the existence of traces of these impurities will result in the corrosion of the engine and other metallic parts (Obaideen *et al.,*2022). They can also lead to the blockage of gas pipelines like in the case of SiO_2 which is formed from siloxanes (i.e., an organic silicon compound related to detergent and lubricant existent in MSW). This thereby incurs additional costs required for either maintenance or replacement of the damaged parts. Also, biogas contains a large fraction of carbon dioxide, which decreases the energy density of the biogas fuel. This is another constraint associated with biogas composition as the removal of carbon dioxide through mechanization as a means of increasing energy density requires extra equipment and energy (Xue *et al*.,2020; Wang *et al*.,2020). Ammonia formed from the degradation of nitrogen-containing compounds such as protein is a very corrosive gas which forms NOx when burned (Braganca *et al.,*2020). The formation of NO_x has a severe greenhouse gas effect capable of contributing to global warming (Braganca et al.,2020; Deublein & Steinhauser.,2011). VOCs are corrosive and have a bad smell with some of them having a toxic effect on the environment (Braganca et al.,2020). H_2S and other sulfur compounds are corrosive and capable of blocking the catalyst's active sites.

Aside from the challenges associated with the management and application of the end products of AD (biogas and digestate), the AD system has some technological challenges which affect the transformation of digestate into valuable materials such as the incomplete biodegradation of the organic matter, the presence of certain complex organic pollutants (e.g., herbicides, fungicides, industrial wastes, hormones (Shargil *et al*., 2015) and the excessive concentration of salts, and pathogenic bacteria (Fecal coliforms 3.60 × 104 –1.06 × 106 CFU g− 1 TS). In addition, data on organic pollutants and other compounds are scarce, and variable due to the heterogeneity of feedstock composition and the type of digestion process.

Another crucial challenge is the wide variety of possible superstructure combinations (feedstocks, process technologies, operating conditions, valuable products, impurities). In addition, the digestate obtained from the AD process has a high-water content. This makes its storage and transportation expensive (Boulamanti *et al.,* 2013; Herbes *et al*., 2020; Silkina *et al*., 2017. Due to the composition of digestate which might contain a wide variety of materials toxic to humans, living organisms, and ecosystems, its disposal is usually a major source of concern (Jomova and Valko, 2011; Silkina *et al*., 2017).

Furthermore, the formation of aromatic hydrocarbons like toluene during the AD of activated sludge is another unpleasant phenomenon of the AD system (Ullah *et al*.,2017; Rasi *et al.*,2007). This is due to the harmful nature of toluene to human health. In addition, the sludge may form a solid layer cake at the bottom of the digester that can cause unsuitable mixing, equipment malfunctioning, increase energy consumption, develop membrane fouling, decrease membrane productivity, and require frequent cleaning (Zacharof and Lovitt, 2014).

Due to the negative impacts of these end products, stringent legislations and standards have been formulated, reviewed periodically, and enforced to address the negative impacts associated with the application of these end products capable of causing pollution, environmental degradation, and the spread of communicable diseases. Though these formulated policies vary across different countries, they have common targets for achieving sustainable development for environmental elements, communicating clear standards and regulations for wastes management, environmental laws to regulate the relevant processes associated with the production of biogas and digestate as well as their corresponding applications (Lamolinara *et al*.,2022).

This can be observed in the EU, which has continuously reviewed its regulations over the years. For example, the action plan for the circular economy (in 2017), for the promotion of waste recycling across the member states to achieve the objective of recycling 70% of the municipal solid waste (MSW) and 5% of landfilling by 2030. In Australia, there are landfill tipping fees, but no legislation to incentivize the process of diverting organic waste from landfills. On the contrary, California's legislation requires municipalities to install AD facilities for the diversion of organic waste (Clarke, 2018). In some countries regulations about digestate and biogas are complex, while in others, the process of reuse is relatively clear (Edwards *et al.,* 2015; Lamolinara *et al.,*2021).

Globally, the UK, Germany, and the USA are countries leading these incentive policy programs. For instance, in the UK, several government programs and supported entities such as the Waste & Resources Action Program (WRAP, 2020), the AD Quality Protocol (ADQP) of the Environment Agency, and the AD, and Bioresource Association (ADBA) have been created. The WRAP program developed the PAS 110 aimed at encouraging the development of the digestate market through the creation of a technical standard industry for producers to check and ensure that digested materials are of consistent quality and fit for the desired purpose. It has also set out the minimum qualifications required for the digestate, separated liquor, and separated fiber which may be used as a fertilizer or soil improver (Pell Frischmann Consultants Ltd, 2012). The ADQP is targeted at providing increased market confidence in the quality of products made from waste, thereby encouraging greater recovery and recycling. This protocol is a set of criteria to produce quality digestate from the AD of material that is biodegradable waste, including the whole digestate, the separated fiber fraction, and the separated liquor. Producers and users are not obliged to comply with the quality protocols. Failure to comply with it will lead to wastage of the quality digestate from AD and waste management regulations will apply to its handling, transport, and application. The ADQP was reviewed by the UK Environment Agency in 2014 to give room for the classification of digestates produced. It also encourages producers to improve their quality (UK Environment Agency, 2014). Finally, the ADBA focuses on consulting services aimed at improving the entire life cycle of biogas and digestate generated at an industrial scale.

The German Renewable Energy Act (REA), in its different versions, has laid the legal foundation for the development of the biogas sector in Germany. This law and regulations associated with it have created advantageous conditions (incentives) for the access of biogas to the electricity markets, as well as measures for a safe investment, and financing of biogas plants (Thran *et al.,* 2020). This is due to the experience, and continuous development of biogas production in Germany (Pfeiffer and Thran, 2018). Efforts are being made to evaluate the impact of this law on the economy, increasing capacity (Scheftelowitz *et al.,* 2018), energy efficiency, and flexible energy supply, as well as the impacts on structural change in agriculture, and investment decisions (Sorda *et al*., 2013). During its progression, this law addressed all the elements of the biogas life cycle, and recent versions include incentives and regulations on the use of digestate as fertilizer. The REA also makes concerted efforts to resolve conflicting objectives that may exist between energy and agricultural legislation (Lamolinara *et al.,*2022).

In the USA, both the US EPA and the USDA have created specific programs to include the Ag-STAR program (EPA-USDA) (US EPA, 2020a) and the Rural Energy for American Program (REAP) (USDA, 2015). The AgSTAR program has developed a set of guides, regulations, information, technologies, and tools to encourage the production of biogas and digestate (US EPA, 2020a; US EPA, 2020b). The USDA's REAP is targeted towards providing financing to small rural businesses to promote the use of renewable energy that includes the production of biogas (USDA, 2020). Other governments and financial incentives applied by the US EPA are Renewable Identification Numbers (RIN) under the Renewable Fuel Standard Program (US EPA, 2020c), Renewable Energy Certificates (RECs) (US EPA, 2020d), and Nutrient Credits (Ross, 2012).

These incentives promote the reduction of organic waste discharged into the environment using it as a bargaining chip for both producers and consumers (Lamolinara *et al*.,2022). The exchange of these credits applies to both fuels and fertilizers in the quest to encourage the utilization of biofuels and biofertilizers. The introduction of quality control of digestate has been a crucial point for its reuse and valorization. Regulation, certification, and quality standards of digestate reinforce the confidence of users in the application of digestate safely and in a manner that respects health, environment, and legal requirements.

The development and implementation of incentive policies and regulations for the sustainable management of organic waste by AD, producing biogas and digestate as valuable products have shown to be highly beneficial. They have also contributed to the advancement of AD technology through the adoption of pretreatment, in vessel cleaning and post treatment techniques amongst others to improve both the performance of AD technology and the quality of the products generated from the AD technology. However, these techniques have been revealed to have various shortcomings such as high capital cost, high energy consumption and sophisticated operating conditions such as maintenance and odor control (Yang *et al.,*2010; Logan & Visvanathan 2019). These shortcomings have hindered their applicability on a full scale. (Weiland 2010; Rongwong *et al.,* 2018). Due to these shortcomings, research on improving the performance of AD systems for the generation of both biogas and digestate has continued to be of interest to the scientific community (Ardolino *et al*.,2021; Kumar Khanal *et al*.,2021). However, more emphasis on the generation of maximum volume of biogas from AD in the quest to meet up with the ever-increasing energy demands of man. This represents a significant gap in knowledge which has prompted the introduction of AI into AD systems for improving the performance of AD in the production of maximum biogas volume as a means of contributing towards filling the significant gap in knowledge.

2.2 Artificial Intelligence

Artificial Intelligence (AI) can be described as computational methods which imitate human mental activities such as thinking, inference, classification, interpretation, estimation, and decision-making (Shehab *et al.,*2020). The history of AI can be traced back to the 4th century B.C when Aristotle invented syllogistic logic, the first formal deductive reasoning system. After the invention of syllogistic logic in the 4th century, an Arab man named Al-Jazari came up with a design in the 13th century which is what is believed to be the first programmable humanoid robot, a boat carrying four mechanical musicians powered by water flow.

Years later, the seeds of modern AI were planted by classical philosophers in their attempt to describe the process of human thinking as the mechanical manipulation of symbols. This work culminated in the invention of a programmable digital computer in the 1940s. This device and the ideas behind it inspired a handful of scientists who began to discuss the possibility of developing an electronic brain. This was later developed to perform difficult tasks in the early years of the 21st century after series of challenges had been encountered in the 1970s and 80s. Having been identified as a tool for performing difficult tasks, AI has constantly been applied in various fields over the years for solving problems and has proven to be successful. It has also been subjected to a series of advancements over the years in the quest to become a better tool/technique.

AI can be divided into four main techniques. These techniques include machine learning (ML), natural language processing (NLP), automation and robotics, machine vision (MV). Due to the limitations of the conventional computational techniques revealed by several research authors to be time consuming and complex in solving analytical problems, AI has been incorporated into almost all fields of study (Yuksel *et al*.,2023). This is due to the capabilities of its modelling techniques in handling multidimensional and noisy data which substantiates the increase of AI application fields. It has several models which have been found highly suitable in various fields of study. In the field of medicine, it is used to perform a better diagnosis where the technologies are used to understand the natural language and respond to the questions asked. In the finance industry, AI is used for collecting personal data and later used to provide financial advice to people. AI has also been found useful in the field of education where the grading system can be automated, and the performance of the students can be assessed based on which the learning process can be improved. Other applications of AI include business to automate the repetitive tasks performed by humans with the help of robotic process automation computer programs like chatbots are used to assist customers in scheduling appointments and ease billing process, smart home devices, security and surveillance, navigation and travel, music etc.

Furthermore, AI-based models have been extensively applied in the field of engineering for tackling engineering problems. In the field of environmental engineering, it has been widely implemented to solve problems related to air pollution, water, and wastewater treatment modelling, simulation of soil remediation and groundwater contamination and solid waste management strategic planning. (Yetilmezsoy *et al*.,2011). They are also capable of solving ill-defined problems, configuring complex mapping, and predicting results (Yetilmezsoy *et al.,*2011). Examples of AI-based models include artificial neural network (ANN), fuzzy logic (FL), expert system and genetic algorithms (GA). Each of these AI models serves a specific function. For example, ANN models can train data for regression and classification. Expert systems such as fuzzy logic can acquire human cognitive and reasoning skills in addition to possessing knowledge base. These systems have a simple linguistic syntax which is proficient in managing complex operations and qualitative attributes (Yetilmezsoy *et al*.,2011). Evolutionary algorithms such as genetic algorithms (GA) adopt the concept of natural selection to obtain optimum results by selecting the best fit data capable of handling unforeseen conditions (Kalogirou;2003a).

Generally, AI-based models offer an alternative effective approach compared to the conventional models and have continued to gain significant attention globally in the scientific community following its discovery. The effectiveness of AI models can be attributed mainly to its ease of use, high prediction accuracy in analysing large amounts of datasets within a short time and a significant reduction of manpower and resource consumption in unnecessary repetitive experiment. Another advantage of AI-based algorithms over the conventional techniques is they do not require multi-disciplinary knowledge related to biokinetics, microbiome, and heat/mass transfer. In addition, the avoidance of model recalibration if trained based on extensive datasets has made it a preferable technique than the conventional techniques. These attributes of AI explain why it has continued to benefit many different industries over the conventional models in the world of today. Hence, the rationale behind its adoption for achieving the purpose of this research study.

2.2.1 Application of AI in AD systems

AI has been applied in AD systems over the years for several purposes as presented in table 2.2 below. Some of these purposes include real-time monitoring, process prediction, process control and stability parameters as well as real-time monitoring (Andrade Cruz *et al*.,2022; Wang *et al.,*2022). The advances in the application of AI in AD for these purposes have been made possible through the use of machine learning (ML) models developed based on AI. The use of ML models in AD is due to its emergence as a data-driven technique independent of complex interactions used in mathematical models (Wang *et al*., 2020).

The ML method is entirely based on either readily available online data or historical recordings of the process (Kazemi *et al*., 2021). It involves three main stages, namely 1. The training stage which involves feeding the algorithm with a training dataset to let the model learn the unseen patterns in the data. 2. The validation stage where a different dataset is used to increase the model's performance by fine-tuning the classifier's hyperparameters; and 3. The testing stage where a different sample of data is used to determine the model's final accuracy (Cruz *et al.,*2022).

ML are grouped into two different types. These two different types include black-box ML and explainable ML. The explainable ML approach attempts to provide a deeper understanding of the functional dependence of the output variables on the input variables. The black-box ML approach on the other hand, is an entirely empirical data-driven modelling technique which does not include phenomenological information on AD. It has two types of modelling techniques: (a) regression (e.g., Neural Network (NN), gaussian process regression (GPR), linear regression, logistic regression, ridge regression, lasso regression, polynomial regression, Bayesian linear regression, etc.) and (b) classification (e.g., SVM, KNN, logistic regression, naive bayes (NB), etc.) (Asgari *et al*., 2021a).

A regression model predicts output variables (e.g., biogas yield) based on numerical (e.g., total solids (TS) or categorical (e.g., reactor type) predictor variables of AD processes. On the other hand, a classification model developed for AD processes is mostly used for the detection of faults or anomalies in an AD reactor such as process inhibition due to VFA accumulation. The explainability of the results obtained from black-box ML models are enabled by approaches such as correlation analysis, feature importance assessment, partial dependence analysis, etc.

Following the emergence of ML models, several ML models have been successfully developed and applied in AD systems for modeling the nonlinear and complex relationships of the AD process (Alejo *et al*., 2018; Zareei and Khodaei, 2017). These models have also been developed to predict biogas yield, process stability parameters and effluent quality indicators (Cruz *et al.,*2022).

For instance, Tufaner and Demirci (2020) employed ANN models in predicting biogas production under controlled laboratory-scale experiments. This was achieved using different input parameters such as reactor fill ratio, influent pH, effluent pH, influent alkalinity, effluent alkalinity, organic loading rate, effluent chemical oxygen demand, effluent total suspended solids, effluent suspended solids, and effluent volatile suspended solids. De Clercq *et al*.,2020 employed RF and extreme gradient boosting-XGBoost in the prediction of methane production from an industrial-scale plant using feedstock composition as the model input parameters.

Similarly, in a pilot-scale study, biogas production from cow manure co-digested with maize straw was predicted and optimised using an ANFIS model under the influencing parameters of C/N ratio, various total solid content and stirring intensity of substrates (Zareei and Khodaei, 2017). Alejo *et al* (2018) investigated the performance of SVM in forecasting the concentration of total ammonia nitrogen (TAN) produced during the AD of a complex substrate. The forecasting of TAN concentration was achieved using TAN, VS, COD_{inf} and TS as the model input parameters. Senol (2021) carried out investigations on the AD of sewage sludge under various ultrasonic pretreatment conditions and predicted its methane yield with ANN, modified logistic model and modified Gompertz model using different sonication times, SE_A values and incremental soluble COD values as input variables. Kazemi *et al* (2021) employed extreme learning machine (ELM) and ensemble neural network (ENN) to monitor and detect random faults in AD processes using pH, pressure, CO_2 and ammonia as the input variables. Seo *et al* (2021) applied RNN in predicting biogas from dry AD of food waste.

The prediction of biogas using RNN was based on SRT, soluble COD, total VFA, total ammonia and free ammonia. Offie *et al.,* (2022) developed an RNN model for real-time monitoring of an AD plant with the aim of improving the operation of an AD plant in producing biogas. This was achieved using different waste compositions of feed and water as the input parameters. Li *et al* (2022) developed an integrated gradient boosting decision tree (GBR) model for predicting methane yield and methane content in AD-generated biogas. This was achieved using different feedstock properties (food waste, manure, algae, and biomass waste), total VFA, methane and microbes as the model input parameters.

Recent studies by Gupta *et al* (2023) revealed that DT-based ML models have been previously developed to predict critical parameters in AD process such as CH_4 yield and CH_4 composition in biogas. For instance, prior research work by Kazemi *et al* (2020b) used an ensemble approach to predict the transient VFA accumulation in AD reactors which is highly detrimental to biogas production. The prediction of transient VFA accumulation in AD reactors was achieved using 13 process variables. These variables include Effluent COD, Effluent Alkalinity, Influent TSS, Effluent TSS, Effluent Ph, Effluent BOD, Gas flow, methane mole fraction, carbon dioxide mole fraction, hydrogen mole fraction, pressure, effluent ammonia, and influent flow.

AI has also been applied in AD systems for optimisation purposes using optimisation algorithms. These algorithms are combined with ML models either as a pre-processor or post-processor. In the pre-processing applications, the optimisation solvers (e.g., GA and ACO) have been used to select the most influential process variables for developing an FNN model that predicted biogas yields (Beltramo *et al*., 2019). The research work indicated that the addition of GA- or ACO-based feature selection to the FNN model reduced the dimensionality of the problem by eliminating superfluous features. This resulted in a reduction of model overfitting. It also improved the accuracy of FNN by 6.2% In another instance, PSO- or GA-based optimisation algorithms have been used downstream to ML models (mainly FNN) for maximising the yield of biogas or CH_4 produced by AD plants (Alrawashdeh *et al.,* 2022; Asadi & McPhedran, 2021; Zaied *et al*., 2020). Integration of FNN with a GA-based multi-objective optimisation framework has also been attempted to determine the pareto frontier (trade-off diagram) between biogas production and effluent COD, which revealed that maximizing the first variable inevitably minimized the latter (Huang *et al.,* 2016).

Recently, Alrowais *et al* (2023) successfully slime mold algorithm (SMA) integrated into a time-series model and a partially connected RNN model for calculating the optimal structure of the RNN model and the optimal value of its parameters such as the optimal number of neurons in the hidden layers, number of feedback connections, activation functions and connection weights.

The successful application of ML models in AD systems has attracted a lot of researchers in the discipline over the years. This is due to its ability to manage complex multivariate data, predict nonlinear connections and handle missing data. In addition, ML online web tools and waste mapping have the potential to enhance the AD plant operator's analytical capabilities, decision making, and planning (De Clercq *et al.*, 2019).

Despite the ability of ML algorithms to manage complex multivariate data, predict nonlinear connections, and handle missing data, selecting the best ML algorithm for performing a certain task in AD systems is highly critical. This is to enable the best results to be achieved using the ML algorithm (Alzubi *et al*., 2018).

2.2.2 Challenges and Knowledge Gap of AI applications in AD

Over the years, significant efforts have been made in the application of ML models developed based on AI in AD systems with the aim of improving AD processes in achieving better output. However, despite these significant efforts made towards developing ML models based on AI for improving AD processes over the years, the development of ML and its application in AD systems is still in its initial stage (Gupta *et al.,*2023). This is because majority of the works done treat ML modelling of AD as a "black box approach" with limited (or zero) physical understanding of process phenomena. This approach poses different challenges.

First and foremost, black-box models are mostly developed based on experimental data obtained from prototypical laboratory scales or pilot-scale reactors, which have led to limited generalizability for industrial scale development (Jia *et al*., 2022). The resultant effect of this has made it potentially problematic for the extrapolation of these models to predict a full-scale system. Secondly, most of the developed ML models applied to AD processes have been observed to lack the presentation of SHAP, permutation feature importance and partial dependence quantification (Gupta *et al*.,2023). These metrics have been revealed to be highly vital for understanding the correlation and variational relationship between predictors (input) and predicted (output) variables. In cases when comparison is done with several types of ML algorithms for automatic optimal algorithm selection, it was observed that most of the research done previously was carried out in terms of their predictive accuracy.

Thirdly, most of the ML models developed for AD processes are either based on metagenomics data or operational parameters. Examples of unifying metagenomic data and operation parameters are rare, for which the generalizability of the model is compromised. Moreover, despite the promising potential of GPRs for model uncertainty quantification, their application in ML modelling of AD processes is still limited as it has not been extensively used.

These challenges associated with the application of ML models in AD systems represent some of the current knowledge gaps associated with the application of AI in AD systems which this research study tends to fill as a means of contributing to knowledge.

For example, the development of the AI-based models adopted for the purpose of this research study was achieved using real data collected from a micro-AD plant operating in an urban area. The adoption of this approach is quite innovative as it not only tested the effectiveness of the developed models in achieving the aims of this research study, but it also explored the feasibility of implementing the developed models on a real case scenario thereby promoting the widespread application of AI on industrial AD systems.

Also, the use of GPR as one of the WLDM for developing the time series ensemble-based model was another innovative approach adopted in this research study. Though it was not applied for the purpose of uncertainty quantification, but its application as a base model for predicting different classes of biogas and as one of the WLDM for first predicting different classes of biogas and then combined with other WLDMs to form the ensemble model to improve the prediction accuracy of biogas using a real case scenario explored the potential capability of the GPR model in the modelling of AD processes.

Moreover, most of the AI applications in AD systems have been carried out for water and wastewater treatment plants as well as for agricultural and livestock waste where AD have been extensively applied for sludge treatment and biogas production respectively. However, the practical implementation of AI models in real AD systems for organic municipal waste is a relatively recent development (Offie *et al*., 2023).

Furthermore, the introduction of user-friendly frameworks such as optimal weekly operation patterns of the different feed variables are rare as most of the previous research works carried out on AD systems focused mainly on the optimisation of biogas using different algorithms to generate maximum volume of biogas from AD systems. Thus, the introduction of optimal weekly operation pattern for the different feed variables was a significant attempt to advance the application of AI in AD systems.

Table 2.2: Further applications of machine learning applications in AD systems

AI model	Purpose of application	Input Variable	Output Variable	Optimal Model	R^2	RMSE	References
ANN	Simulation and optimisation of biogas production from Russeifah biogas plant	Slaughterhouse waste, restaurants, fruits, vegetables, and dairy markets	Biogas	GA-ANN	0.87		Qdais.,2010
ANN	To predict the rate of both biogas and methane production from a pilot scale mesophilic UASB reactor To determine the effectiveness of ANN in predicting biogas and methane production by comparing it with other non-linear regression models.	Volumetric organic loading rate (OLR), influent and effluent pH, operating temperature, influent and effluent alkalinity, effluent chemical oxygen demand (COD), and volatile fatty acid (VFA) concentrations	Biogas Methane	-	0.94 0.92	0.062 0.065	Yetilmezsoy *et al.*,2013
RNN	To predict the rate of biogas production from an AD plant	Flow rate, dry matter, volatile solids, pH, and temperature	Rate of biogas production	-	-	0.004	Dhussa *et al.*, 2014
FFNN	To evaluate methane yield from biogas in a laboratory scale anaerobic reactor.	Moisture content, pH, total volatile solids, and volatile fatty acids	Biogas yield	ANN-BP	0.73		Nair *et al.*, 2016
FFNN	To optimise biogas from food waste.	Substrate mix percentage, plant pH level, digestion period and digester temperature	Biogas yield	BPNN	-	0.000006	Palaniswarmy *et al.*,2017
FFNN	To model and optimise the oil, refinery wastewater and chicken manure.	Total ammonia nitrogen, free ammonia nitrogen, total content of volatile fatty acids, pH, acetic acid, propionic acid, butyric acid, and temperature	Biogas production Biomethane content	ANN-LM	0.88 0.51		Mehryar *et al.*,2017
FFNN	Modelling of biogas is produced from cattle manure with the co-digestion of different organic wastes.	Working days, influent chemical oxygen demand, Influent pH, influent alkalinity, influent ammonia, total phosphorus, hydraulic retention time, waste adding ratio, and pretreatment and reactor number.	Biogas yield		0.75		Tufaner *et al.*,2017
FFNN	To predict the rate of biogas produced from potato starch processing wastewater.	Ammonium, COD, pH, alkalinity, total Kjeldahl nitrogen, total phosphorus, volatile fatty acid, and hydraulic retention time	Biogas rate, Methane rate.	ANN	0.98		Antwi et al., 2017

AI model	Purpose of application	Input Variable	Output Variable	Optimal Model	R^2	RMSE	References
	To predict the methane rate from potato starch processing wastewater.						
FNN	Modelling and optimisation of specific biogas production from mixed ligno-cellulosic co-substrates under mesophilic and thermophilic conditions.	Feedstock composition HRT, temperature	Biogas yield	-	0.99	43 ml/gVS	Ghatak & Ghatak (2018)
FFNN	Prediction of biogas production using ACO and GA input features selection method	Concentration of VFA, total solids, volatile solids acid detergent fibre, acid detergent lignin, neutral detergent fibre, ammonium nitrogen, hydraulic retention time and organic loading rate	Rate of biogas production		0.9	0.0624	Beltramo et al.,2019
FFNN	To predict biogas from chemically treated co-digested agricultural waste. To optimize cumulative methane from the produced biogas.	Composition of the substrate, operating temperature, dose of the $NaHCO_3$, and hydraulic retention time (HRT)	Cumulative methane		0.99		Almomani *et al.,* (2020)
RF, FNN, ELM, C-SVM, GP	Online monitoring of volatile fatty acids in AD processes	pH, TAN, pressure, TS, COD, ALK, gas flow, mole	VFA	GP	0.99		Kazemi *et al.* (2020b)
LSTM, DA-LSTM, DA-LSTM, VSN	To predict and improve the production of biogas from an anaerobic co-digestion process.	Sludge inflow and outflow, temperature, SRT. VS/TS, BOD, COD, SS, TN, TP	Biogas production	DA-LSTM-VSN	0.76		Jeong *et al* (2021)
GLMNET, RF, XGBOOST FNN, KNN, C-SVM	To investigate the feasibility of six ML algorithms in predicting methane yield.	Feedstock composition, operational conditions, and genomic data	Methane yield	RF	0.82	40 ml/gVS	Long *et al* (2021)
ANN, MG	To predict the cumulative biogas (CBY) and methane yield (CMY) from the anaerobic digestion of several organic wastes	volatile solid to total solid ratio (VS/TS), carbon content, carbon to nitrogen ratio (C/N), carbon content and digestion time		GA-ANN	0.9996 0.9998	0.0045 0.0046	Mougari *et al*., (2021)

AI model	Purpose of application	Input Variable	Output Variable	Optimal Model	R^2	RMSE	References
XGBOOST , C-SVM, RNN, RF	Improving the prediction accuracy of methane yield using different machine learning models.	pH, alkalinity, COD removal efficiency, VFA	CH_4 yield	RNN	0.97	23ml/ gCOD	Park *et al* (2021)
RF, XGBOOST , C-SVM, LSTM, RNN	To improve the process stability of bio electrochemical anaerobic digestion processes (BEAD) using five different ML models	OLR, pH, alkalinity, VFA, and COD removal efficiency	CH_4 yield	RNN	0.97	0.025	Cheon *et al* (2022)
Linear regression, GLMNET, KNN, C-SVM, DT, RF, XGBOOST	To predict the possible impacts of wide range temperature fluctuations on the process of AD	Temperature, pressure, feed, volume, and nutrient solution usage	CH_4 yield	C-SVM	0.85	-	Cinar *et al* (2022)
RF, XGBOOST, FNN	To explore various machine learning algorithms for predicting the changes in the abundance of antibiotic resistance genes in anaerobic digestion	Operating mode, feedstock pre-treatment, additives, temperature, and HRT	Relative abundance of ARG and MGE	FNN	0.79	-	Haffiez *et al.* (2022)
C-SVM, RF, AdaBoost, XGBOOST	To predict the concentration of total volatile fatty acids in multiple full-scale food waste anaerobic digestion systems with different machine learning models and feature analysis	Feedstock composition, VS, TS, HRT, pH, ALK, COD	VFA	XGBOOST	0.64		Choi *et al* (2022)
Hybrid machine learning based on LSTM and RF	To improve the prediction of biogas generation output by redefining the key input parameters	Kitchen waste, starchy waste, human faecal waste, bagasse waste, pig manure waste, leachate waste, and chicken litter	Biogas yield		-	15.26	Chiu *et al.*,2022
RNN-SMA, RNN	Compare the effects of thermophilic & mesophilic anaerobic co-digestion for sustainable biogas production using an experimental and RNN model study.	Temperature, Digester time, Carbon-Nitrogen ratio	Biogas production	-	0.99 0.99	0.007 0.00052	Alrowais *et al* 2023

3 Chapter 3. Methodology

This chapter presents the different procedural steps adopted for addressing some of the technical limitations of the AD technology. The different procedural steps adopted in this research study also tend to address some of the economic challenges/limitations associated with the AD technology. Details of this will be provided in section 3.1. Thus, these steps are explained in a manner that enables its broad adoption in similar AD projects. To this effect, the case study used for the proposed methodology is explicitly presented in section 3.1. The different procedural steps adopted are targeted towards achieving the aims and objectives of this research study. The procedural steps adopted are also targeted towards addressing the research questions raised in chapter 1 of this research study which are centred mainly on improving organic waste management using AD, facilitating the implementation of circular economy, and contributing towards achieving the various SDGs linked to AD. Firstly, a new AI-based framework developed for the prediction and optimisation of biogas obtained from a micro-AD plant based on data driven models. The proposed methodology adopted for developing a new AI-based framework is presented in Figure 3.1. The new AI-based framework presented in Figure 3.1 comprises of three main stages namely data collection and preparation, model development and performance assessment. Secondly, a novel approach for the development of a time series ensemble model for improving the accuracy of biogas predictions from the micro-AD plant is presented. The development of the ensemble model is based on different weak learning data mining (WLDM) models. The WLDM models are also data driven. The proposed methodology adopted for developing the ensemble model is presented in Figure 3.3. Similar to the new AI-based framework, the ensemble model is also comprised of three main stages. These stages include data collection and preparation, feature extraction and selection, model development and performance assessment. The stages of both frameworks are commonly used for developing most data-driven environmental models (Piadeh *et al.,*2022). Both frameworks are mainly used as the core tools for estimating and optimising biogas generation based on the feed data collected over preceding days. Each of the steps outlined in both frameworks were conducted using the MATLAB 2021b software which provides functions both for the estimation and optimisation of the AD performance.

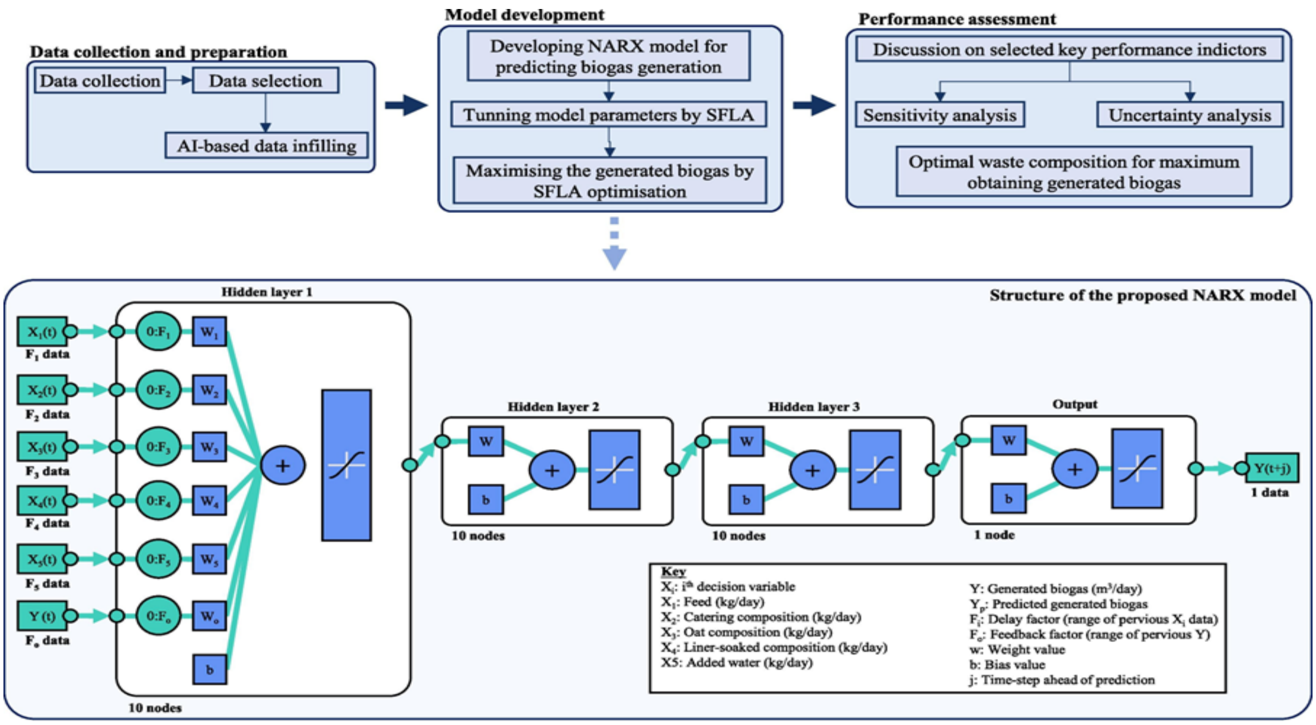


Figure 3.1. AI-based framework for the operation of the micro anaerobic digestion plant

3.1 Description of the case study

The purpose of this research study was achieved using a micro-AD plant located in Camley Street Natural Park Central London, United Kingdom. The micro-AD plant used as the pilot study for this research was a wet system which ran continuously for a period of 310 days at an average temperature of 35.7˚C (i.e., under mesophilic conditions). It was mainly used for treating commercial organic waste produced within the locality. Further details of the micro-AD plant are provided in section 3.2. The selection of this specific micro-AD plant for this research was due to the operational challenges associated with the micro-AD plant initially selected for this research study. The effect of this made it difficult for the efficient collection of data required to develop the AI-based models. Also, the lack of laboratory equipment necessitated the need for a pilot plant to be investigated in the first place. Hence, the rationale behind the collaboration with the operators of the micro-AD plant located in Camley Central London to adopt their micro-AD plant for achieving the purpose of this research study.

The adoption of this micro-AD plant was targeted towards addressing some of the technical and economic challenges affecting the implementation of AD technology for managing organic waste. The effect of these challenges has hindered the implementation of circular economy and the achievement of the various SDGs linked to AD. Also, the adoption of a micro-AD plant for managing organic waste has been revealed to have offer great support in addressing organic waste management issues especially in urban/municipal areas (Walker *et al*.,2017). Moreover, micro-AD plants have been identified to have great potential in addressing one of the major economic challenges associated with conventional AD plants (Walker *et al*.,2017). This economic challenge includes the reduction of transportation costs through the establishment of a micro-AD plant in urban/municipal areas where organic waste is generated in large quantities. This makes it easy for the generated organic waste to be sent to plants unlike the conventional AD plants mostly located at city/community outskirts.

In addition, the establishment of micro-AD plants in urban/municipal areas offer promising potential for community involvement compared to conventional AD plants which are located at city/community outskirts which are non-resident areas. This helps to promote peace in communities, especially in communities around the world where peace is non-existent as the establishment of micro-AD plants in these communities for biogas generation will help to bring people together for achieving a common goal beneficial to the development of the community. Another benefit of the micro-AD plant is its ability to foster circular economy by means of creating a ‘biorefinery’ that will dispose of local waste, utilize its energy potential, and produce a natural fertilizer that can be used in urban agriculture, horticulture, and hydroponics. Despite these benefits associated with the micro-AD plant, circular economy is yet to be fully implemented thereby hindering the realization of the various SDGs. This is due to the technical limitations associated with AD which have hindered the performance of AD in producing maximum volume of biogas sufficient to satisfy the ever-increasing energy demands of the people. Also, the application of micro-AD technology is not very common in many parts of the world as most research studies previously carried out on AD have focused mainly on conventional AD technology (Sawatdeenarunat *et al.,*2015).

Hence, the development of a new AI-based framework and ensemble-based model framework in improving the performance of AD for maximum biogas generation is not only targeted towards further addressing some of the identified drawbacks but also making micro-AD technology a more familiar and accessible technology globally for treating organic wastes. The effect of this approach could potentially increase the uptake of micro-AD technology by increasing understanding of the field and capturing feedstocks from sources that are out of the catchment area of larger plants.

3.2 Data collection and preparation

This stage entails the collection of raw data from an AD plant, analysis of the raw data collected, imputation for infilling missing data using some data-mining-based techniques and the selection of relevant data for the development of the RNN model. The data used in the development of the AI-based model was collected from a micro-AD plant. The micro-AD plant was designed as a pilot study by a consortium of researchers and companies. As stated in section 3.1, the micro-AD plant is located in Camley Street Natural Park Central London, United Kingdom (UK) with the schematic diagram shown in Figure 3.2. The micro-AD plant was built within the grounds of the Camley Street Natural Park in London, UK (Walker *et al.,*2017). The site was used to convert the locally produced commercial organic waste collected by cargo bicycles into biogas for cooking, heating, and electricity purposes (Walker *et al.,*2017). It had a manual screener for removing impurities from collected food waste, a pre-feed tank with a volume of 0.65 cubic meters, a grit/inert container for storing the grits and contaminants removed from the waste and a feed pump as illustrated in figure 3.2. It also had a main anaerobic digester of volume 2 cubic meters containing an automated mechanical mixer (Methanogen UK Ltd., UK) and heater by an internal water heat exchanger. Other main components of the micro-AD plant as shown in figure 3.2 include the hydrogen sulphide scrubber filled with activated carbon pellets for controlling odour and pollution from hydrogen sulphide, floating gasometer for biogas storage, digestate sedimentation tank which had a volume of 0.46 cubic meters for storing the digestate obtained after the AD process. It also a digestate liquor storage tank of 0.2 cubic meters. The micro-AD plant was monitored for a period of 310 days (i.e., approximately 10 months during which the operational parameters, biological stability, and energy requirements of the micro-AD plant were evaluated.

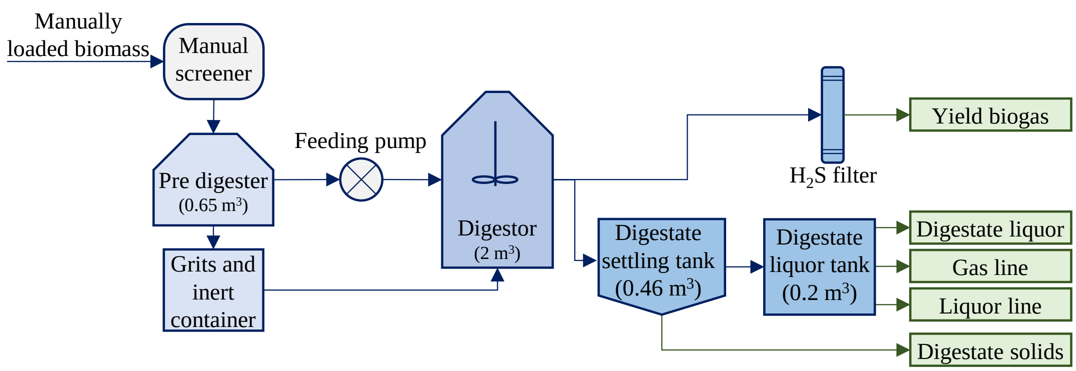


Figure 3.2. Schematic diagram of the micro-AD plant used in this study.

The data collected from the micro-AD plant include temperature, pH, volatile solids, total solids, feed into the main digester, feed composition into the pre-feed tank. The feed composition was made up of apple, catering and coffee, coffee, digestate, green waste, oats, soaked peanuts and muesli, tea, tea leaves, tea bags, oil, soaked muesli, soaked liners, and catering. Other sets of data collected are the water added to the pre-feed tank and the volume of biogas generation. The feed into either the pre-feed tank or the main digester was usually done every few days when both feed amounts and biogas volume in the storage were recorded. The data collected enabled the calculation of the total feed and water added to the AD plant over its operational period of 310 days, hydraulic retention time (HRT), total biogas production as well as the average overall, specific, and volumetric biogas production. These data collected represent the factors influencing the performance of the micro-AD plant used as the pilot study for producing biogas. It was observed that out of the monitoring period of 310 days of the micro-AD plant, there were days when no feed was added to either the pre-feed tank or the main digester and no volume of biogas generated from the plant was recorded while daily continuous data for both feed and biogas are necessary for developing a time-series ANN model that takes lag days into consideration.

Furthermore, some days were observed to have missing output data (i.e., there was feed but there was no reading for volume of biogas generation). The effect of this can hinder the accuracy of the developed model of the micro-AD plant especially for the prediction of the biogas volume generated and the optimisation of biogas thereafter. Hence, to overcome this barrier some data-mining techniques were first analysed in this research study for estimating the missing data. This was to determine the most suitable one for infilling the missing data. It is worth noting that the missing data in this research study refers to the absence of biogas readings in two types: (1) data samples with feed values available (input) but no reading for biogas generation (output); and (2) data samples with feed value equal to zero but no reading for biogas generation. Therefore, the entire dataset was first divided into two groups of data with feeding inclusive and data without feeding. Some data mining techniques were then tested to identify the relationship between the feed data and the generated biogas for data groups with feeding data. Out of those data mining techniques, the one which had the least range of data fluctuations was regarded as the best data mining technique for infilling the missing data. Hence, the best data mining technique was selected for infilling the missing data of the first group (i.e., data with feed values but no biogas values) as presented in figure 4.2 under the results section. The second type of missing data (i.e., data where feeding is zero and biogas data was not recorded) were infilled based on the linear regression of the remaining total biogas data read. The data mining techniques explored in this research study include Random Forest (RF), K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), Kriging, Feed Forward Neural Network (FFNN) and Linear Regression (LR). Sequel to the infilling of missing data using data mining technique, the RNN model was developed as described in section 3.2. Sequel to this, a sensitivity analysis was carried out for each of the operational feed variables to determine their correlation and impact on the volume of biogas generation.

Based on the initial cross-correlation analysis of all input variables demonstrated as Fig. 4.1 and Table 4.2 presented in the results section, the daily feed into the digester, the water added to the digester showed significant correlation and corresponding impact on the biogas volume generated. In addition, out of various waste compositions, only oats, soaked liners, and catering were selected whereas other waste compositions were negligible as they were observed to have no significant correlation and hence no meaningful impact on the volume of the biogas generated from the micro-AD plant.

Generally, the production of biogas from an AD plant is influenced by multiple factors which are interconnected. These factors include feedstock composition, environmental parameters (temperature, pH, moisture content and HRT) and organic loading rate (OLR) as stated in the literature review section. However, some of the environmental parameters such as temperature, pH and HRT were excluded from the analysis for estimating biogas generation. OLR was also excluded from the analysis for estimating biogas generation. These parameters were excluded following the cross-correlation analysis conducted for each of the AD parameters which revealed that the excluded parameters were relatively constant during the operation of the micro-AD plant. This was indicated by the lag times observed for each of the operational parameters as presented in figure 4.2. In addition, the volatile solids and total solids which were measured were also observed to be relatively constant during the operation. Hence, these operational parameters were also excluded from the analysis for estimating biogas generation.

3.3 Model Development for RNN/NARX

The development of the RNN/NARX model for the purpose of this research study, was achieved in three stages as illustrated in figure 3.1. These stages include 1. Developing the model for predicting biogas generated from the micro-AD plant. 2. Tuning the model parameters using shuffled frog leap algorithm (SFLA) 3. Optimising the biogas generated used SFLA. The RNN/NARX model was developed with three hidden 10-neuron layers with the architecture shown in the figure 3.1. This model was developed based on the selected input variables of the micro-AD plant. These input variables include the actual/estimated daily feed added to the main digester denoted as X_1 in figure 3.1, the feed composition which comprised of catering denoted as X_2, oats denoted as X_3, and soaked liners denoted as X_4 added to the pre-feed tank (i.e., the top three highly correlated variables with biogas generation), the water added to the pre-feed tank denoted as (X_5), and the volume of biogas generation which represented the output variable denoted as Y. The model was then set for training, validation, and testing. The model settings are as follows: Levenberg-Marquardt method used for the training process, mean square error (MSE) as the indicator for evaluating the performance of the model and 6 epochs (iterations) were adjusted for training failure. The database used for the model development was divided into three parts. These parts include 70% for training, 15% for validation and 15% for testing. The trained model was then used to predict the biogas generation (Y_p) from the micro-AD plant in the case study. Sequel to this, the trained model was tuned using an optimisation method as the NARX model needed lag time specification in days (known as delay factor F_i), i.e., range of input variables for previous timesteps to use for estimation of 1 timestep ahead (Y_{t+1}) based on input data (decision variables). This was done to obtain the most accurate output data (biogas generation).

The employment of an optimisation method was used to determine the optimal lag time for each decision variable as presented in the results section. This was also aimed at obtaining the most accurate output data i.e., biogas generation. In addition, the optimisation method enabled the maximum volume of biogas generated to be obtained.

Furthermore, the optimisation method showed the best distribution of each of the decision variables required to obtain maximum volume of biogas. Based on these, the optimisation model was developed using shuffled frog leaping algorithm (SFLA). SFLA is a memetic meta-heuristic and nature-based algorithm highly useful in solving combinational optimisation problems (Eusuff & Lansey.,2006). It has the ability to search in both local and global search space where each lag time represents one frog (Bui *et al.*, 2020).

In this research study, each frog, i.e., decision variable, represents a lag time to find the minimum root mean square error (RMSE) and the highest Normalized Nash-Sutcliffe Efficiency (NNSE) for this optimisation approach. Hence, 4 trials for exploration and 4 trials for exploitation were set for each iteration of optimisation, with the stopping criteria being set to an improvement of less than 1%. Each of the six decision variables (i.e., F_0-F_5 illustrated in Figure 4.3 under the results section) is an integer value ranging between 0 to 10 due to the results of cross-correlation analysis on inputs provided in Figure 4.1 under the results section. Through this approach, the delay factor (range of previous X_i data) for each input data/decision variable can be determined. It was then used again to specify the required weights for the daily feeds added to the main digester, daily feed compositions and the water added to the pre-feed tank to maximise the output (i.e., maximum volume of biogas generation from the micro-AD plant) for each of the days in a cyclic period of feeds. It is pertinent to note that the cyclic period was based on the (lag time) delay factor specified in the first optimisation model. While the stopping criteria and trials were set similarly, each NARX input for each day were selected as the decision variable. To simulate the real operation and put a cap for the feeds/water added to the plant, constraints were defined based on the historic operation of the micro-AD plant. These constraints are as follows: (1) maximum feed equals to 80 kg every 4 days. Where the 4 days in this context was based on the cyclic period of 5 days (i.e., four days of input data and biogas generation on the fifth day) specified as a result of the optimisation model for the largest lag time (as presented in the results section); (2) total weight of feed and all pre-feeding compositions should be equal during the optimisation; (3) added water is limited to 30% of total feed weight, (4) all decision variables need to be either zero or positive values.

3.4 Performance assessment of the RNN Model

Two metrics i.e., RMSE and NNSE were used in this research study as the performance indicators for evaluating the performance of the developed RNN model. These two metrics were selected as the performance indicators based on previous case studies where both metrics have been extensively applied and proven to be reliable indicators for evaluating the accuracy of a developed model. The RMSE measured the error between the predicted biogas volume and the actual biogas volume measured from the micro-AD plant. The NNSE on the other hand estimated the variation between the predicted biogas volume and the measured biogas volume.

Based on these two metrics, the effectiveness of the RNN/NARX model in achieving the aims and objectives of this study was determined. To further confirm the effectiveness of the developed model, sensitivity analysis was carried out for each of the operational input parameters. The essence of the sensitivity analysis was to determine the correlation and impact of each of the input parameters on the predicted output (volume of biogas generated). This was done by removing one input parameter and running the model afterwards with NNSE and RMSE both under observation. When conducting the sensitivity analysis, the optimal waste composition was taken into consideration where the impact of each waste composition was analysed and evaluated to determine the significant impact of each waste composition on the generation of biogas. The impact of each waste composition on the generation of biogas was then compared with the operator’s analysis. This measure was taken to further confirm the accuracy of the developed model. In addition, an uncertainty analysis was also conducted to show how the relative accuracy changes when running the model with dataset reduction. The essence of this was to prevent overconfident predictions which can be harmful when applied in real-life scenarios.

Thus, the uncertainty analysis was conducted by reducing the dataset at 10% interval with the relative accuracy of the developed model obtained for each corresponding reduction in the dataset observed. Further analyses of biogas generation for the different decision variables were also conducted on the feed, water, waste composition, and the impact of the distribution of the optimal values of the pre-feed composition variables on biogas generation for oat, catering and liner data. The essence of these analyses was to further evaluate its importance in the performance of the developed RNN model for improving the performance of the AD plant in generating biogas. The results of these further analyses carried out on the developed RNN model will be presented and discussed extensively in the results section.

3.5 Methodology for the Development of the Ensemble Model

The proposed framework for the development of the ensemble-based model for the real-time operation of AD in the production of maximum biogas volume was carried out in three main stages as presented in figure 3.3 below. These stages include: 1. Data collection and preparation/Data Acquisition 2. Model development and 3. Performance assessment. To ensure the applicability of the proposed framework to scenarios with limited data availability, only data obtained from a single micro-AD plant commonly used for treating food wastes within an urban area were used for the purpose of maintaining simplicity. Also, the single micro-AD plant used for this purpose was the same with that used for developing the RNN/NARX model. The data collected from the single micro-AD plant were numerical and time-series data. These data include soaked oats, soaked liner, tea or coffee residuals, and cupboards, which were fed to the pre-digester after separation and screening through grits or other means. Additionally, the volume of water added to the pre-digester, feeding rates to the main digester, and the amount of biogas produced. These data were collected at varying intervals, ranging from a few minutes to daily Similar to the development of the RNN model, the data collected had missing values which were cleaned based on recommendations from Offie *et al.,* 2023.

The numerical and time-series data collected from the micro-AD plant were transformed into features and selected based on the further explanation provided in Section 3.5. The relevant features were extracted from the data collected which were subsequently transformed into group features.

Each of the group features were also categorized into different classes namely zero, low, medium, and high. Further details of this will be explained in section 3.5. These features were used for the development of the weak learning data mining models (WLDM) explained in section 3.6. The weak learning data mining models were developed and stored in a data warehouse along with their key performance indicators (KPIs), which serves as the foundation for constructing the proposed ensemble model. Details of this will be provided in Section 3.6. Following the construction of the ensemble model, rigorous testing was conducted on real-time unseen data to evaluate its performance under real-world conditions. The outcomes of this testing and a detailed analysis of the results are presented and discussed in the results section. This analysis provides an in depth understanding of the model's effectiveness and its potential for real-time optimisation in practical scenarios.

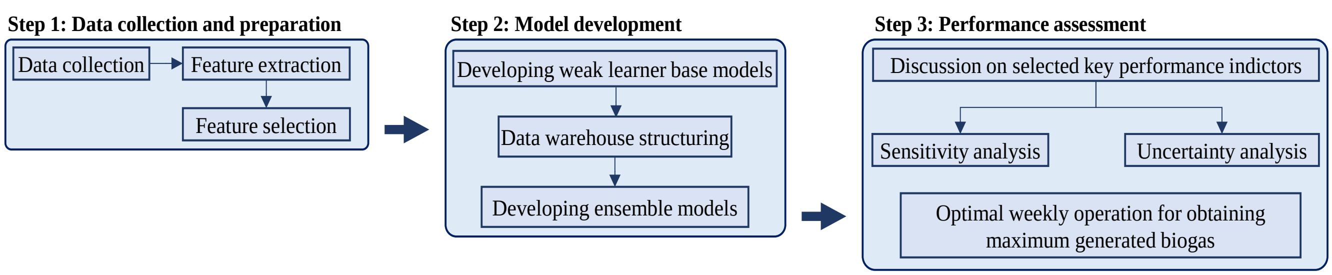


Figure 3.3: Proposed framework for the development of the ensemble model

3.6 Data Collection and Preparation

This stage involves the extraction and selection of features relevant for the development of the model. However, this procedure was conducted sequel to the collection of raw data from the micro-AD plant, analysis and infilling of missing data observed in the dataset using different data mining techniques as explained in section 3.1. Similar to the procedural steps for development of the RNN model, the temperature, HRT, VS, TS, and pH values were observed to be constant throughout the AD operational process as the values for each of the parameters were within a specific range. Hence these parameters were not taken into consideration when developing the ensemble model. As earlier stated, the relevant features were extracted from the collected dataset and classified into group features. This is presented in Table 3.1 below. These group features include feed, catering composition, oats composition, liner composition, water, and biogas feature as shown in table 3.1. Each of these group features contained 7 features which represented the 7 days of weekly operation. Each of these group features were further grouped into four different classes for feed, catering composition, oats composition, soaked liner composition and the water added to the digester. These classes include, 0 (zero), L(low), M(medium), and H(high). In the case of the biogas feature, it was grouped into three different classes namely L(low), M(medium), and H(high). This is also described in table 3.2 below. The classification of both the feed and biogas features were based on the operational conditions of the micro-AD plant and view of the expertise. On the other hand, the water feature was classified based on the clustering method. Following this, the extracted features presented in table 3.1 were then refined using three established techniques namely principal component analysis (PCA), partial least squares (PLS), and sequential sensitivity analysis (SSA). These techniques are widely accepted as prerequisite steps for the identification of the key variables enhancing classification performance and reducing computation times (Masahiko *et al.,* 2019).

Principal Component Analysis (PCA) is a sophisticated statistical method widely applied in data science especially for reducing the dimensionality of data sets (Jollife & Cadima 2016). It focusses mainly on the transformation of complex high-dimensional data into a simpler, more manageable form while preserving significant information. This is mainly achieved through the extraction of principal components from the datasets. The PCA technique, otherwise known as dimensionality reduction technique, has become integral in various fields of discipline ranging from finance to genomics, where the management and interpretation of vast amounts of data is a common challenge (Jollife & Cadima 2016). It has also been widely used in wastewater treatment plants for process monitoring and in AD plants for diagnosing the state of anaerobic digestion (Li & Yan 2019). In addition, PCA has been applied at the preliminary stages of data analysis for uncovering hidden trends in the data. Moreso, it is useful in multivariate data analysis, dealing with observations of multiple interrelated variables.

PLS is a technique used to estimate linear relationships between dependent and independent variables. It shows the direct effect of independent variables on the dependent variables. In addition, it can reduce the dimensionality of correlated variables, modelling the underlying and shared information of the variables. PLS has been widely applied in anaerobic digestion processes for monitoring the process parameters, biogas prediction and the simulation of the best operational condition for the generation of maximum biogas volume (Awhangbo *et al.*,2020). It has also been applied as a multivariate data analysis technique which combines the methodologies of regression and linear analysis (Boulesteix & Strimmer 2006). It is widely employed in the different fields of study where handling unobservable or hidden variables is vital. As such, it has been found to be highly useful in the field of business, management and accounting, social sciences, economics, finance, environmental sciences, medicine, and health professions amongst others.

Sequential Sensitivity Analysis (SSA) is a technique used to determine the impact of each input/decision variable on the output variable (biogas). It is an integral part of anaerobic digestion modelling as it provides information on the most influential parameters on the model output (Barahmand & Samarakoon., 2022). SSA also reduces the amount of estimation required. SSA plays a vital role in modelling AD systems which helps in the operation of a biogas production unit.

Hence, sequential sensitivity analysis (SSA) was carried out by removing one feature at a time and measuring the accuracy difference of the developed WLDM models. Based on the analysis, the key features were then selected and stored in a data warehouse for developing and testing the WLDM models. This analysis was conducted for each of the group features as presented in the results section.

Table 3.1: Potential features extracted for developing weak learning data mining models.

Group Feature	Data Unit	Description	Class
Feed	**kg/day**	The weight of organic material sent to the digester for decomposition. It was measured in kg.	0: zero feeding L: feeding less than 20 kg/day M: feeding between 20 and 40 kg/day H: feeding more than 40 kg/day
Catering	**% daily feed**	The composition of food waste produced in commercial kitchens, canteens, and restaurants.	0: zero ratio in daily composition L: less than 16% in daily composition M: between 16% and 50% in daily composition H: more than 50% in daily composition
Oat	**% daily feed**	The composition of organic wastes containing oat grains	0: zero ratio in daily composition L: less than 9% in daily composition M: between 9% and 45% in daily composition H: more than 45% in daily composition
Liner	**% daily feed**	The composition of organic wastes containing liners soaked in liquid.	0: zero ratio in daily composition L: less than 7% in daily composition M: between 7% and 63% in daily composition H: more than 63% in daily composition
Water	**% daily feed**	The amount of moisture content added to the digester during its operations.	0: zero ratio in daily composition L: less than 5% in daily composition M: between 16% and 40% in daily composition H: more than 40% in daily composition
Biogas	**m^3/day**	The end product of the entire anaerobic digestion process. It was measured in volumes.	L: less than 1 m^3/day M: between 1 and 4 m^3/day H: more than 4 m^3/day

0: Zero L: Low M: Medium H: high
***: Each group feature contains 7 features representing 7 days of weekly operation**
Classification is provided based on operational practices and desired industrial goals. Range of each class are provided by the analysis on raw data and using k-mean classification model

3.7 Model Development

The development of an ensemble model involved the combination of multiple WLDMs to create a more robust and accurate biogas prediction model. For the purpose of this research study, the stacking method was selected. The selection of this method for achieving the purpose of this research study was on the basis of its previous applications in different fields of study, where it has proven to be well suited for creating ensemble models from multiple WLDM models. The stacking method has the ability to reduce both bias and variance (Gupta *et al.,*2023) The ability of the stacking method to reduce both bias and variance distinguishes it from the bagging and boosting method which can reduce only one of both (Chauhan 2021). This particular attribute of the stacking method helps to improve its accuracy in the prediction of biogas as it combines the predictions of different WLDM models otherwise known as base models to obtain a final prediction (Chauhan 2021). It also makes the model more flexible to be applied to any machine learning algorithm. In addition, it makes the model more robust than other ensemble models thereby making it less susceptible to overfitting, The development of WLDMs involved six different techniques namely: decision trees (DT), k-nearest neighbour (KNN), gaussian process regression (GPR), support vector machine (SVM), naïve Bayes (NB), discriminant analysis (DA). These specific models were selected based on their widespread application and recognized potential in previous anaerobic digestion processes, where they have been used for various purposes (Cruz *et al.,* 2022; Gupta *et al.,* 2023; Khan *et al.,* 2023). Each model was developed using MATLAB 2022b and optimised through automatic hyperparameter optimisation, aiming to minimise the five-fold cross-validation loss over thirty iterations. Further details on the optimisation process are presented in Figure A3 in the Appendix. In addition, error bias was mitigation using the 5-fold cross-validation method. The dataset was divided into three distinct portions for training and testing the models. Specifically, 60% of the dataset was allocated for training the individual WLDM models, 20% of the dataset was reserved for testing the performance of these models while the remaining 20% of the dataset was set aside for evaluating the proposed ensemble model. To ensure a balanced and equal representation of the databases, the group features were randomly distributed across the training, validation, and testing databases (See Table A3 in the Appendix).

Subsequently, the built WLDM models were stored in a model library and their KPIs are stored in data cube. The KPIs of the developed models in predicting biogas was assessed using the confusion matrix concept as a statistical classification technique (Grandini *et al.,* 2020; Tharwat, 2021). This technique involved mapping the predicted biogas classes (i.e., low, medium, high) onto the confusion matrix. Using this mapping of the confusion matrix, this study employed two main KPIs of true positive rate (TPR) i.e., ratio of correct prediction of i^{th} class of yielded biogas and true negative rate (TNR) i.e., ratio of correct rejection for situation in which yielded biogas is not ith class. These two KPIs are determined for each class of low (class 1), medium (class 2), and high (class 3) yielded biogas. TPR and TNR rate are determined based on Equations 1 and 2. Model library and data cube is integrated as a data warehouse that used for developing ensemble model.

$$\mathrm{TPR_I}\ (\%) = \frac{\mathrm{TP_i}}{\mathrm{n_i}} \times 100 \qquad \textbf{Equation 1}$$

$$I\ (\%) = \frac{\mathrm{TN_i}}{\mathrm{n_i}} \times 100 \qquad \textbf{Equation 2}$$

where TPR_i is the TPR of ith class, TNR_i is the TNR of ith class, TP_i is the number of correct ith class prediction, TN_i is the number of correct rejections of non- ith class prediction, and n_i is the total number of measured ith class.

The ensemble model was developed by combining the developed WLDMs to create a more robust and accurate prediction model. As earlier stated, the stacking method was selected due to its widespread application in different fields of study. Most importantly, it has been successfully applied in previous AD systems (Mukasine *et al*.,2024) This method involved the training of all WLDMs on the same set of training data. The WLDMs were blended afterwards using a decision tree framework inspired by bucket of models' method, as shown in Figure 3.4. To determine the class of yielded biogas, a set of WLDM models are adjusted. These models are then selected based on their higher performance in each key performance indicators (KPIs) previously stored in data cube. For example, as shown in Figure 3.4, group 1 models are the models in which TPR_1 was recorded in the range of acceptable (e.g., DA, DT, and NB model in Figure 1). Here this rate was selected as 70% based on recommendations of Cruz *et al.* (2022), Gupta *et al.* (2023), and Khan *et al.* (2023). The predicted class by these models were blended by voting techniques. The predicted classes i.e., F_1, F_2, and F_3 were fed into a decision tree framework to determine the final prediction. This framework operates under specific conditions to identify the most appropriate predicted class. In scenarios where a single predicted class aligns with the selected group, that particular class is chosen as the final prediction. To illustrate this process, consider an example where group 3 predicts a high value for F_3 (F_3=H), and F_2 is predicted to be anything other than medium (F_2=L or H, but not M), and similarly, F_1 is predicted to be anything other than low (F_1=M or H, but not L). Under these conditions, the correct predicted class would be H, following the fifth left branch of the decision tree in Figure 3.4. On the other hand, if all models predict their respective classes, the model with the highest average of TPR takes precedence. This is demonstrated in the first right branch of the decision tree presented in Figure 3.4. This approach to ensure that the model with the highest degree of accuracy/correct estimation is selected when all models agree on their predictions. In cases where none of the models can accurately predict their respective classes, the final decision is made by assessing the overall performance of these models using the highest average of TPR and TNR. This criterion is represented by the first left branch of the decision tree in Figure 3.4. When faced with situations where two models strongly advocate for their respective classes and are unable to reach a consensus, the model with the highest Score value (S_{ij}), as determined based on equation 3, is selected as the final decision.

$$S_{ij} = \mathrm{TPR}_{\,i\,for\,group\,i} + \mathrm{TNR}_{\mathrm{i\ for\ group\ j}} \qquad \textbf{Equation 3}$$

where Sij is the determined score, i and j are the two selected groups.

For this purpose, two scores are determined. As an example, let us consider a situation where F_3 and F_2 are predicted as H and M, respectively, and F_1 is predicted as not L. In this scenario, the first score is calculated as the summation of the TPR of group #3 and the TNR of group #1 (S_{31} in the second right branch of the decision tree in Figure 3.4). The other score is computed as the summation of the TPR of group #2 and the TNR of group #1 (S_{21}). To make the final decision, the two scores are compared. If S_{31} is greater than S_{21}, F_3 is selected as the final prediction. Otherwise, F_2 is chosen. This approach effectively evaluates the capability of true prediction of the two-group model (consisting of groups 1 and 2 in this example) based on the TNR rate of the other model group (group 3). By employing this decision tree framework, the system systematically determines the final prediction in cases involving multiple predicted classes from distinct groups. This method ensures a structured and reliable approach to arrive at the most suitable prediction based on the aligned groups' predictions.

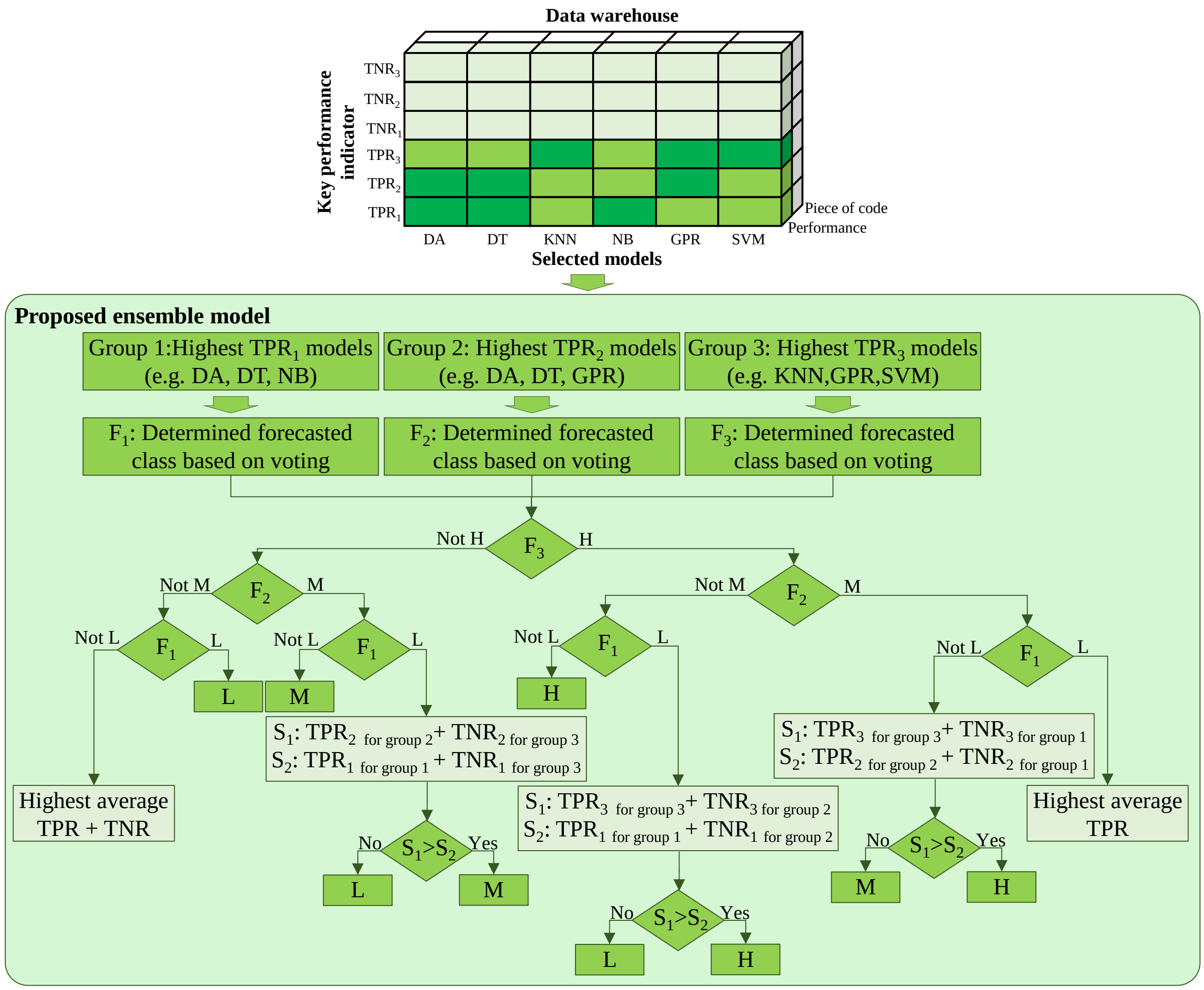


Figure 3.4: **Data warehouse and proposed ensemble model**

3.8 Real Time Weekly Operation

Sequel to the development of the ensemble model, the developed model was then optimised using SFLA. This was to determine the optimum condition for the weekly operation of micro-AD plant for maximum biogas generation. The developed model was optimised for 7-days ahead to determine the optimum condition for the weekly operation of the micro-AD plant. To achieve this objective, a set of constraint rules was integrated into the process of selecting the optimal scenario. These constraint rules served as guiding principles to determine the best course of action, considering a range of factors. The key constraint rules were established based on the view and technical expertise of the micro-AD operator. The constraints serve as general recommendations for other similar AD projects which seek to apply these principles to optimize patterns either in their research or practical projects. The following key constraint rules were then established: (1) minimising input loads to ensure that the highest possible yielded biogas was obtained for each unit of added materials, optimising resource utilisation, (2) minimising input days to mitigate the operational costs associated with material handling, transportation, and processing, (3) minimising added water load aligning with the goal of conserving water resources and mitigating associated energy costs, resulting in more sustainable and efficient operation, (4) minimising feeding days for cost savings arise from frequency of operational activities and associated resource consumption. By integrating these constraint rules, the optimisation framework strived towards striking a balance between the maximisation of biogas production, minimisation of resource inputs, and the optimisation of operational costs. This comprehensive approach ensured that the chosen scenarios were in accordance with achieving both environmental sustainability and economic efficiency goals which have sustainable societal benefits. By understanding and following this pattern, operators can effectively optimise their operations to maximise biogas yield. The proposed pattern was evaluated against conventional AD operation for a period of one month.

3.9 Performance assessment of the developed WLDM & Ensemble models

Several benchmark models were developed to facilitate a comparative analysis with the developed ensemble model. These models are intended to further enhance predictive performance and provide a comprehensive evaluation of the different techniques used. The implementation of these benchmark and ensemble models serves as a valuable reference point for evaluating the effectiveness of the developed models in improving the biogas prediction accuracy. These benchmark models include: (1) hard voting stacked model determining final class based on the majority class label predicted by the individual WLDM models, (2) soft voting stacked model considering the probabilities or confidence scores assigned by each WLDM model for each class and blending them to make the final prediction. Furthermore, optimised stacking models are created by combining the best performing developed WLDM models including (1) the ensemble of the best performing WLDM in TPR_1 i.e., low based, (2) the ensemble of the best performing WLDM in TPR_2 i.e., medium based, and (3) ensemble of the best performing WLDM in in TPR_3 i.e., high based. Moreover, to ensure the generalization and comprehensiveness of the proposed model, other blending methods such as bootstrap aggregating (Bagging) and boosting were also selected including the optimised versions of (1) RF, and (2) Subspace of developed NB, (3) XGBoost, (4) Gentle boos of developed DA model, (5), Random under sampling and boosting (RUS Boost) of GPR model. The shuffled frog leaping algorithm (SFLA) optimisation technique, along with the classification and optimisation toolboxes of MATLAB 2022a, were employed to identify the best type of documented various models of stacking, bagging, and boosting. During the optimisation, the number of learners varied from 1 to 500, the learning rate ranged from 0.001 to 1, the maximum number of splits varied from 1 to 18618, and the classification error improvement threshold was set at 0.01%. To assess the performance of ensemble models, in addition to TPR and TNR, Accuracy or correct prediction of all classes (ACC), false positive ratio (FPR) i.e., portion of abnormal prediction, overestimation rate, and underestimation rate were used as the performance assessment metrics. These metrics are determined based on Equations 3-6.

$$\text{ACC}(\%) = \frac{\sum \text{TP}_i + \text{TN}_i}{\sum n_i} \times 100$$ **Equation 3**

$$\text{FPR}_i = \frac{\text{FP}_i}{\text{TN}_i + \text{FP}_i}$$ **Equation 4**

$$\text{Overestimation}(\%) = \frac{\sum \text{FP}_i}{\sum n_i} \times 100$$ **Equation 5**

$$\text{Underestimation}(\%) = \frac{\sum \text{FN}_i}{\sum n_i} \times 100$$ **Equation 6**

Where FPR_i is the FPR of the i^{th} class, FP_i is the portion of the situation in which i^{th} class is predicted as higher yielded biogas, FN_i is the portion of the situation in which i^{th} class is predicted as lower yielded biogas.

Furthermore, sequential sensitivity analysis (SSA) was carried out to determine the impact of each feature on the generated biogas. To achieve this, one feature was removed at a time and the accuracy difference of the developed WLDM models was measured. This method serves the purpose of gaining insights into the impact of each feature on the proposed ensemble model. Uncertainty analysis was also carried out to show changes in the relative accuracy with corresponding reductions in the dataset. The results of both the sequential sensitivity analysis (SSA) and uncertainty analysis will be presented and discussed under the results section.

3.10 Summary

The proposed methodology for this research study explored three different novel approaches in achieving the aims and objectives of this research study. These three different novel approaches include 1. The size of the dataset used for the development of both RNN and Ensemble-based model 2. Application of SFLA for the optimisation of the developed models.3. Incorporation of time-series concepts into the ensemble model. These three novel approaches will be discussed in detail as follows.

The size of the dataset used for model development: The size of the dataset used for the development of both the AI-based and Ensemble model was obtained from an AD plant which was observed to be limited as it had an operational period of 310 days (i.e., an operational period of approximately 10 months). The utilisation of this size of dataset for the development of both the RNN and Ensemble model is quite rare especially in the development of the RNN model (i.e., a neural network model). This is due to the fact that series of research studies conducted previously on AD systems using RNN have demonstrated its ability to effectively predict biogas production and other AD output variables such as methane production amongst others using datasets having longer periods of operation ranging from a period of at least 1 year to a period of 8 years (Dhussa *et al.,*2014; Wang *et al*.,2021). In addition, the development of neural network models using datasets of large sizes can be attributed mainly to the fact that the use of limited dataset for model development has frequently been observed to be less effective in improving the performance of AD systems compared with traditional machine learning models (Feng *et al.,*2019). To this effect, datasets of large sizes are generally desired for the development of neural network models to ensure a higher level of model accuracy (Seo *et al*.,2021).Moreover, the size of the dataset previously used in developing different ML models such as SVM, DT, KNN etc for predicting biogas and other AD outputs have shown analysis over longer periods of operation ranging from 2- 8 years demonstrating satisfactory performances (Liu *et al*.,2022; Wang *et al.,* 2021). The utilisation of dataset having limited size (i.e., 310 days) for the development of both RNN and the ensemble model for achieving the aims and objectives of this research study, further tested the feasibility of the developed models. It also provided a useful insight on the ability of both models to improve the performance of an AD plant with limited period of analysis. This will enable the right decisions to be made in the future by the operator on the best approach to adopt in order to further improve the effectiveness of the AD system.

<u>The application of SFLA as the optimisation tool</u>: SFLA was used as the optimisation tool for the developed RNN and Ensemble models to improve biogas production with the aim of obtaining maximum volume of biogas from the micro-AD plant. It was also used to determine the optimal daily feeding pattern that can yield maximum volume of biogas from the micro-AD plant. The application of SFLA for the optimisation of biogas production from AD is also a novel approach. Though SFLA has commonly been applied in different aspects of civil engineering such as water resources management for solving optimisation problems (Guo *et al.*,2020), construction/project scheduling to assist decision-makers in the identification of the best Pareto solution for time-cost-resource trade-off (TCRTO) problems under the constraint of precedence, resource availability, and on-site peak electricity power load (Tao *et al.,*2019) and pier maintenance. Its application in for the optimisation of estimated biogas produced from the operation of AD is highly innovative as previous research studies have employed different optimisation algorithms in the optimisation of estimated biogas production from AD. These different optimisation algorithms include Genetic Algorithm (GA), Hybrid Bayesian Optimisation (HBO) method and Adaptive Neuro-Fuzzy Inference System (ANFIS) (Kana et al.,2012; Sadoune et al.,2023; Zareei &Khodaei (2017). Other optimisation algorithms applied in the optimisation of estimated biogas production from AD include Ant Colony Optimisation (ACO), Response Surface Methodology (RSM) and Seagull Optimisation Algorithm (SOA) (Beltramo *et al*.,2016; Dahunsi *et al*.,2017; Abdel daiem *et al*.,2022). Furthermore, Particle Swarm Optimisation (PSO) algorithm has also been applied in the optimisation of biogas (Zeinolabedini *et al*.,2023). The application of SFLA for achieving the purpose of this research study further tested the effectiveness of SFLA in the optimisation of estimated biogas from AD operations. It will also serve as a decision-making tool for future researchers who intend to adopt optimisation algorithms to further improve biogas production.

Incorporation of Time-Series into the Ensemble Model: The incorporation of time-series into the ensemble model is yet another novel approach adopted for achieving the purpose of this research study. This is because previous research studies carried out on ensemble models developed using different ML models have either focussed on investigating the feasibility of the ensemble models developed from different ML models used in predicting biogas and other AD outputs or predicting the correlation between operational parameters and biogas production quantities obtained from a real-scale AD plant. (Long *et al*.,2021; Xu *et al.,*2021; Oznur & Yildirim 2023).

Additionally, the feasibility of the developed ensemble models has been investigated using different performance indicators such as coefficient of determination for their evaluation (Wang *et al.,*2023; Li *et al.*,2022). The incorporation of time-series into the ensemble model represents a significant advancement aimed at enhancing the capability of the developed ensemble model to improve the performance of the AD plant in generating maximum biogas volume. Also, the introduction of a user-friendly weekly operation pattern enables easy implementation by AD operators.

Furthermore, this approach will make it relatively simple and practical thus enabling its straightforward application in various industrial settings. This will allow operators who may not have extensive technical expertise in advanced modelling techniques to easily understand and interpret the input variables and the corresponding output classes. Moreover, dealing with different volumes and numbers in a practical setting can be highly challenging and cumbersome where these models may simplify the decision-making process by providing clear indications of the system's operational state or the class to which inputs or outputs belong.

These three novel approaches have contributed towards filling a significant gap in knowledge. They also helped to address some of the technical and economic limitations associated with the real-life application of AD. Through this means, the novel approaches have provided answers to the research questions raised in the introduction of this research study. However, some key operational and environmental parameters of the AD plant were excluded when developing both models. These parameters include temperature, pH, HRT, OLR, TS and VS. The exclusion of these parameters was on the basis of the cross-correlation analysis which were observed to be constant during the operation of the AD plant thereby having no significant correlation on the volume of biogas compared to the feed, composition variables (i.e., oat, catering and liner) and water added. Further details will be presented in 4.5 under the results section.

Although these parameters influence the performance of the AD process in biogas production, they are unsuitable for the detection of abnormal conditions (Kazemi *et al.,*2021). For instance, a decrease in biogas production or pH implies that instabilities have already occurred in the process. Also, the OLR and HRT values are dependent on the decision of the plant's operator (Kazemi *et al.,*2021). Normally, these parameters change when there is a change in feed composition or due to instability of the process. Moreover, early indicators of process imbalance such as gas composition measurement (CH_4, CO_2, H_2), redox potential, alkalinity and VFA concentration can provide apriori indications of the process imbalance, they do not give direct information as regards to the exact cause of process imbalance (Boe *et al.,* 2010; Dixon *et al*., 2007).

Table 3.2: Summary of the Novel Approaches adopted in this Research Study.

Novel Approach	Description of Novelty	Justification of the Novel Approach	References
Size of the Dataset used for the development of both Models	The utilisation of this size of dataset for the development of both the RNN and Ensemble model is quite novel	Previous research studies conducted on AD systems using RNN and other ML models for the Ensemble model have demonstrated their ability to effectively predict biogas production and other AD output variables using datasets having longer periods of operation ranging from a period of at least 1 year to a period of 8 years	(Dhussa *et al* 2014; Liu et al., 2022 ; Wang *et al* (2021)
SFLA optimisation algorithm	The adoption of SFLA as the optimisation algorithm method in obtaining maximum volume of biogas from AD is also novel	SFLA has commonly been applied in different aspects of civil engineering such as water resources management for solving optimisation problems, construction/project scheduling, time-cost-resource trade-off (TCRTO) problems and pier maintenance	(Guo *et al.*,2021; Tao *et al.*,2019)
Incorporation of time-series into the developed Ensemble Model	This represents a significant advancement aimed at enhancing the capability of the developed ensemble model. It enables its easy implementation by AD operators. straightforward application in various industrial settings.	Previously developed ensemble models had either focussed on investigating the feasibility of the ensemble models developed or focussed on predicting the correlation between operational parameters and biogas production quantities obtained from a real-scale AD plant	(Long *et al.*,2021; Xu *et al.*,2021; Oznur & Yildirim 2023)

4 Chapter 4. Results and Discussion

The results of the proposed methodology adopted for the development of both the new AI-based framework and the time-series ensemble-based model in predicting and optimising biogas from a micro-AD plant is presented and extensively discussed in this chapter. The discussions will be based on the observations made in the results presented for the different analysis carried out for both the new AI-based framework and the Ensemble-based model respectively. The discussions will also focus on how the results obtained can help to address some of the technical limitations of the micro-AD plant as well as the potential impact they have on the environment, society, and economy. This is in accordance with achieving the aims and objectives of this research study. Furthermore, the discussions tend to provide answers to the research questions outlined in the scope of this study using the novel approaches adopted in the proposed methodology.

4.1 Performance Statistics of the Micro-AD plant

Table 4.1 shows the performance statistics of the various operational parameters of the micro-AD plant used as the pilot study for achieving the aims and objectives of this research study. The various operational performance statistics presented in Table 4.1 were obtained using the raw data collected from the micro-AD plant. The raw data collected enabled the calculation of the total feed and water added to the micro-AD plant over its operational period of 310 days. It also enabled the calculation of other parameters used in the development of both models. These parameters include average daily biogas volume and volumetric daily biogas volume. Other parameters such as volatile solids. HRT, average temperature of the digester, average OLR, average biogas methane content etc were also recorded during the operation of the micro-AD plant as presented in Table 4.1. These operational parameters were used for the purpose of comparison with the results of the developed AI-based model. This was to further confirm the effectiveness of the developed models in improving the volume of biogas generated from the micro-AD plant. From table 4.1, it can be observed that the data obtained on the average daily feed amount and average volume of daily biogas produced from the micro-AD plant, was calculated based on the daily biogas data recorded during the operation of the micro-AD plant.

Table 4.1; Key Operational Performance of the Micro-AD plant

Measurements	Value	Unit
Average daily feed amount	14.3	kg/day
Average daily volatile solid added	3.22	kg/day
Average organic loading rate	1.6	kg VS $m^{-3/}$day
Average water added	2.3	kg/day
Average daily biogas production	3.15	m3/day
Volumetric daily biogas production	1.57	m^3 biogas m^{-3} digester day^{-1}
Total mass of food added to the plant	4574	kg
Specific biogas yield	220	m^3 $tonne^{-1}$fresh matter
Specific methane yield	595.5	m^3 CH_4 $tonne^{-1}$ VS
Average biogas methane content	60.6	%
Average hydraulic retention time	127.2	days
Average temperature of the digester	35.7	˚C

4.2 Outcome of the Data Mining Techniques for Infilling the Missing Data

Figure 4.1 shows the performance of the different data mining techniques applied for infilling the missing data observed in the data collected from the micro-AD plant. The performance of each of the different data mining techniques was evaluated using the RMSE of the test data based on the cross-validation method in which all data samples took part in the evaluation of the test set of data (Eghbali *et al.,* 2017). The 6-fold cross-validation method was used in this research study for assessing the performance of each of the data mining techniques employed for infilling the missing data. Based on the results presented, it was observed that the Kriging technique had the least range of fluctuations compared to the other data mining techniques as it demonstrated to have an average RMSE value of 1.23 m^3/day unlike the other techniques which were observed to have RMSE values within the range of 1.25-2.25 m^3/day. This implies that the application of the kriging technique will give rise to a closer range of data compared to the other data mining techniques employed in this research study.

Thus, the Kriging technique was selected and used to obtain the missing biogas values with complete feed values thereby giving rise to more accurate predictions of the biogas produced from the AD plant. This further confirms the effectiveness of the Kriging technique in infilling missing data where other previous studies have not highlighted the challenge of missing data and have in most cases, used simple common techniques such as linear regression and linear interpolation for infilling missing data (Pei *et al.,* 2022; Seo *et al*.,2021). The performance of the kriging technique also confirms its effectiveness over widely known infilling techniques such as KNN and SVM which have proven to be effective. This is especially important because similar to many industrial and real practices, the generated biogas was not measured daily and only 40% of the non-sequential data were recorded. This represents 123 non-sequential data out of the 310 days of operation.

Hence, the application of this infilling technique was highly useful as it was able to address the challenge of missing data which is a major challenge currently being faced in many industrial and real practices. It provided acceptable results used to develop the RNN/NARX model used for accurate prediction and optimization of the generated biogas volume from the micro-AD plant. The acceptable results provided by the kriging technique were also applicable for the development of both the six different WLDM models and the ensemble-based model thereafter.

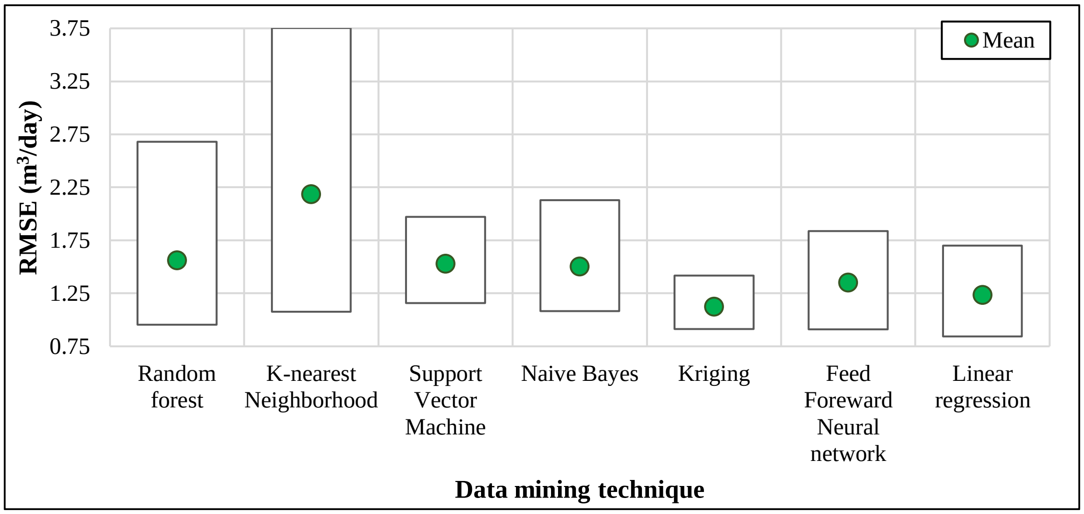


Figure 4.1. Performance of data mining techniques applied for infilling the missing data.

4.3 Cross Correlation Analysis of the Operational Parameters

Table 4.2 and Figure 4.2 presents the results of the cross-correlation analysis carried out on the raw data collected from the micro-AD plant. This analysis was a measure of the operational parameters of the micro-AD plant where the impact of each of the operational parameters on biogas was measured with respect to lag time which indicated the point where the best match between the operational/input parameter and biogas occurred. As stated in the methodology, this analysis was carried out to determine the input parameters which had a significant impact on biogas yield.

From figure 4.2a, it was observed that feed parameter had a correlation coefficient of 0.60 at the 5^{th} lag. This implies that biogas was generated 5 days after the addition of feed into the AD plant. In figures 4.2b and c, the TS and VS parameters were observed to have a correlation coefficient of 0.99 at the 0^{th} lag respectively.

The occurrence of both TS and VS coefficient at the 0^{th} lag indicates that both TS and VS had no significant impact on the biogas yield. This can also be observed in the STD of the non-zero value for both TS and VS presented in Table 4.2 which appeared to be relatively low compared to the other parameters.

Figure 4.2d showed that catering had a correlation coefficient of 0.42 at the 3rd lag. This implies that biogas was generated 3 days after the catering was added into the AD plant. In figure 4.2e, it was observed that oats had a correlation coefficient of 0.31 at the 3rd lag thereby implying that biogas was also generated 3 days after the addition of oats to the AD plant. Similarly in figure 4.2f, liner was observed to have a correlation coefficient of 0.38 at the 3rd lag just like catering and oats. The similarity in the lag time between catering, oats and liner indicates their significant impact in the volume of biogas generated from the AD plant.

Figure 4.2g showed the correlation between water and biogas where water had a coefficient of 0.38 at the 2nd lag. This implies that biogas was generated 2 days after the addition of water to the AD plant. In figure 4.2h, the correlation between the digester temperature and biogas was illustrated where temperature was observed to have a coefficient of 1 at the 0th lag. The occurrence of the coefficient at the 0th lag indicates that temperature has no significant impact on biogas. Hence, the results of the cross-correlation analysis of the raw data presented in table 4.2 and figure 4.2 indicates that the biogas generated from the AD plant was mainly influenced by feed, catering, oats, liner, and water.

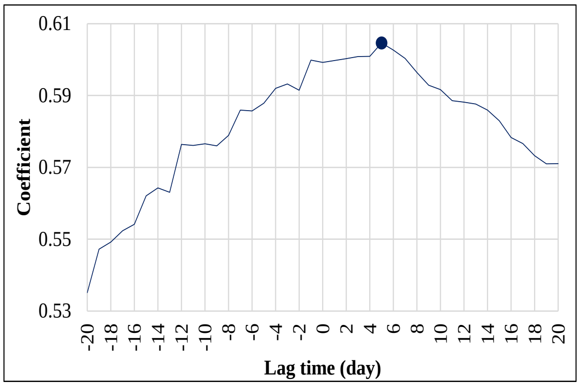
Table 4.2: Initial Cross-Correlation Analysis of the Raw Data

Parameter	**Recorded zero value (%)**	**STD* of non-zero value**	**Cross – correlation****
Feed (Kg/day)	50	28.28±17.33	0.60 at 5th lag
TS (%)	0	0.26±0.04	0.99 at 0th lag
VS (%)	0	0.25±0.04	0.99 at 0th lag
Apple Pomace (Kg/day)	100***	NA+	NA
Catering (Kg/day)	28	32.64±24.40	0.42 at 3rd lag
Coffee (Kg/day)	100	NA	NA
Feedstock (Kg/day)	100	NA	NA
Green waste (Kg/day)	100	NA	NA
Oats (Kg/day)	11	30.26±12.56	0.31 at 3rd lag
Soaked muesli (Kg/day)	100	NA	NA
Soaked liners (Kg/day)	22	5.66±4.12	0.38 at 3rd lag
Soaked peanuts (Kg/day)	100	NA	NA
Tea (Kg/day)	100	NA	NA

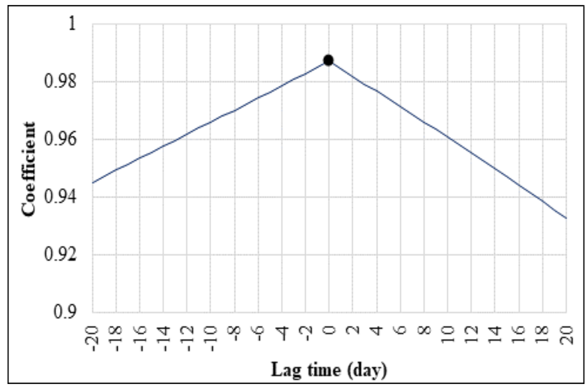
Tea leaves (Kg/day)	100	NA	NA
Tea bags (Kg/day)	100	NA	NA
Oil (Kg/day)	100	NA	NA
Water (Kg/day)	28	3.5±0.71	0.38 at 2nd lag
Biogas production (m^3/day)	0	3.26±1.21	-
Digester temperature (^oC)	0	32.90±0.13	1 at 0th lag

*: Standard deviation **: Cross-correlation between interested parameter data and biogas data

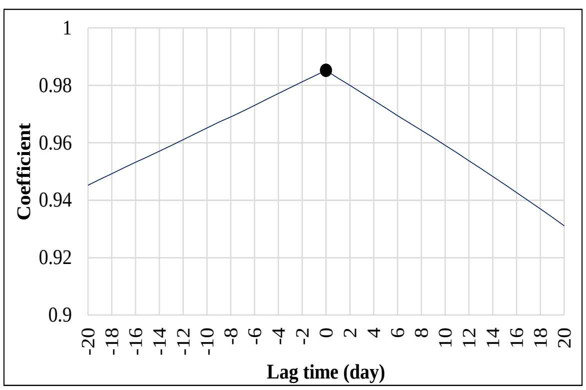
***: Lack of data +: Not applicable



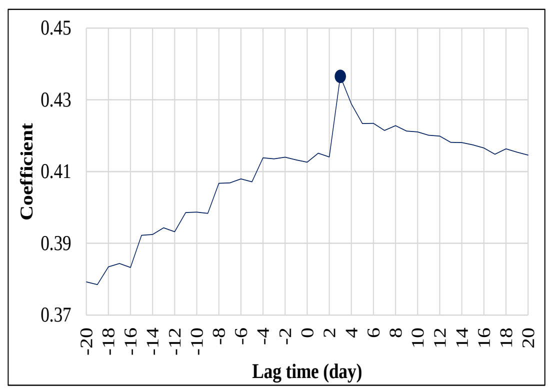
(a)



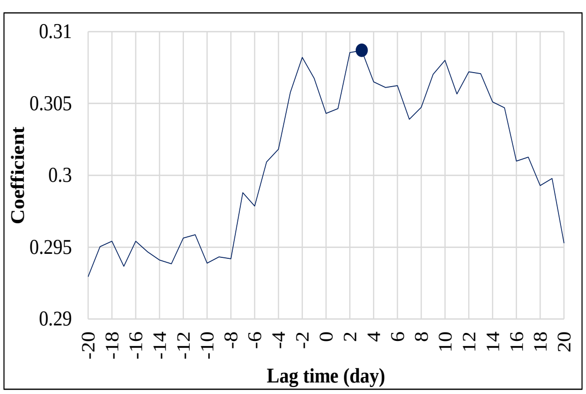
(b)



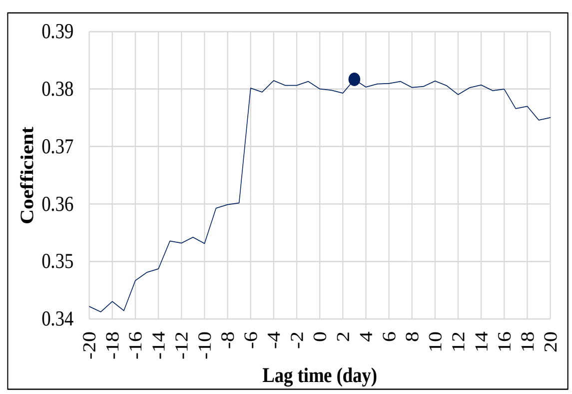
(c)



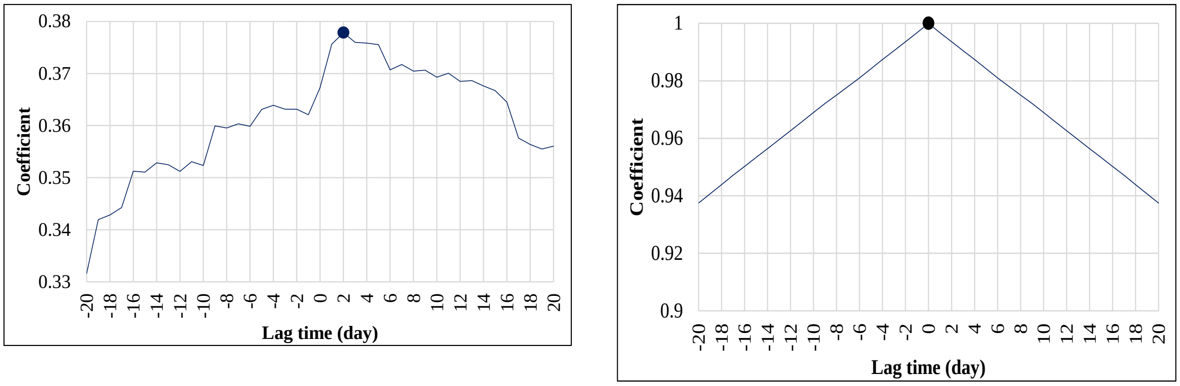
(d)



(e)



(f)

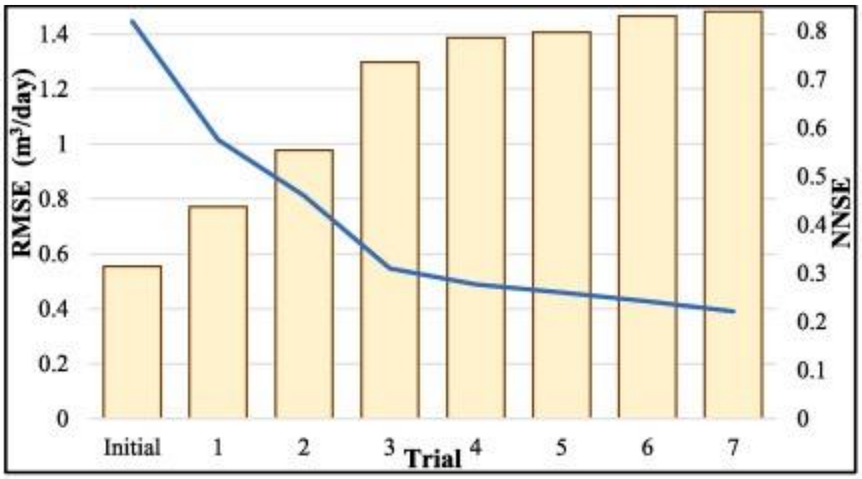


(g) **(h)**

Figure 4.2. Cross-correlation between the yielded biogas and: (a) feed, (b) TS, (c) VS, (d) catering, (e) oat, (f) liner, (g) water, (h) temperature

4.4 Outcome of the developed RNN model tuning and calendar distribution of inputs

Sequel to the infilling of missing data, the RNN model was developed using the datasets and tested afterwards using both RMSE and NNSE described in the methodology as the performance indicators. The optimal number of lag times for each input variable can be seen in Figure 4.3 below. This was obtained after 8 trials as shown in figure 4.3 below. The figure also shows the obtained lag times in each iteration and their corresponding model performance metrics. As it can be seen in the figure, the RMSE was observed to decrease for each trial step ahead while the NNSE increased concurrently. This was expected as the closer the RMSE value is to zero, the more accurate the model is in giving precise predictions of biogas Also, the closer the NNSE value is to one, the model accurate the model is in predicting biogas. Based on the result presented in figure 4.3, it was observed that the longest optimal daily lag time is 5 days for the added water variable, followed by 3 days for feed added to the main digester, and then 1 day for other variables (i.e., catering composition, oat composition, soaked-liner composition, and biogas generation). These values obtained indicate that the generated biogas is influenced by the long term and gradual effects of added water and daily feeding weight whereas the oat, catering and soaked liner waste compositions can immediately impact on biogas for only one day. It was also observed that the yielded biogas is heavily dependent on the calendar distribution of feeding and water in comparison to the weights of added feeding material to the digester. Furthermore, one day lag time in the waste composition indicates that the process of biogas generation is highly impacted by the different rate of compositions which are added to the pre-digester, even if this ratio is not completely similar to the material entered into the digester.



Trial	Optimised number of lag times (days)					
	F_1	F_2	F_3	F_4	F_5	F_o
Initial	5	3	3	3	2	1
1	6	1	2	2	2	1
2	4	2	1	1	3	2
3	7	1	2	1	3	7
4	10	2	1	1	3	11
5	3	1	1	1	5	5
6	3	1	1	1	5	3
7	**3**	**1**	**1**	**1**	**5**	**1**

F_1: Feed F_2: Catering composition F_3: Oat composition F_4: Liner-soaked composition F_5: Added water F_o: Generated Biogas
NNSE — RSME

Figure 4.3. The trend of the SFLA optimisation method to determine optimal daily lag time of input variables.

Moreover, the cross-correlation analysis, illustrated in Fig.4.2, shows the highest correlation co-efficient between biogas generation and the feed added to the main digester occurs in previous 5 days (used as initial for F_1 in Fig. 4.3) whereas optimised lag time for feed is reduced to 3 days (trial 7 in Fig. 4.3). Similarly, daily lag times for catering, oat and liners were observed to decrease from 3 days in initial trial based on the cross-correlation analysis to only 1 day. However, the high correlation of 5 to 3 days before for the added water were initially ignored (number 2 in the initial row for F_5 in Fig. 4.3 vs number 5 in the last row). This can be due to the impact of the combination of input variables on optimal lag times that is shown in the significant improvement of metrics, i.e., RMSE (decrease from 1.4 to 0.4) and NNSE (increase from 0.6 to around 0.9). Although most of the previously developed NARX models recommended using cross-correlation results for developing NARX model (Abdel daiem *et al*., 2022), the difference between initial lag times and final lag times obtained from the SFLA method shows the added value of using optimisation models to fine tune these time-series models. This shows that the applied optimisation method could speed up the modelling while increasing the accuracy as it was observed in the RMSE values which significantly decreased to 0.4 from 1.4 while the NNSE had a significant increase to approximately 0.9 from 0.6. This is very important because although almost all research studies that previously used RNN/NARX model, recommended the use of cross-correlation results for developing the RNN/NARX model (Abdel daiem *et al.,* 2022), the difference between the final delay factors, i.e., obtained days, and initial values experienced an increase in value through the application of the optimisation models to fine tune these time-series models. Thus, the ability of the applied optimisation method to fine tune the RNN model is highly beneficial as it can improve the efficiency of the AD process in the production of biogas thereby, addressing one of the major technical limitations of AD mostly encountered in real industrial practices.

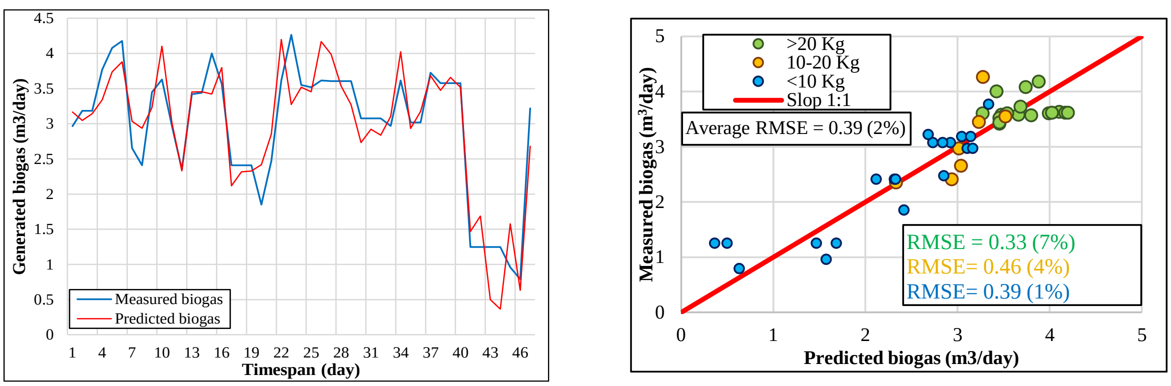
4.5 Performance of the Developed RNN Model

In figure 4.4a, comparisons were made between the measured biogas and the predicted biogas over the test period of 47 days. Based on the figure presented, sudden drops and rises in the volume of the predicted biogas were observed along days 4-10, 16-19, 22-25 and 40-43 respectively. These sudden drops and rises indicate signs of instability in biogas production from AD operation along these days. These signs of instability can be attributed to the exclusion of the temperature parameter in the development of the RNN model as different research studies have revealed that the temperature parameter plays a vital role in ensuring the stability of the AD system which influences biogas production from the AD system (Dela-Rubia *et al.*, 2002; Bouallagui *et al*., 2009b; Riau *et al.,* 2010). Thus, the exclusion of temperature parameter in the development of RNN gave rise to the sudden drops and spikes in the volume of the predicted biogas observed along days 4-10, 16-19, 22-25 and 40-43 respectively. This further confirms the significance of temperature in the production of biogas from AD operations.

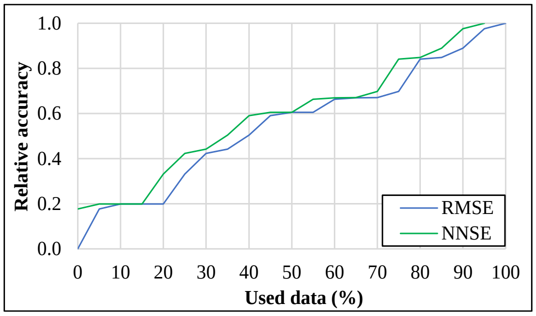
Figure 4.4b also shows the performance assessment of the developed model where the scatter plot of predicted biogas versus corresponding measurements for one day lead time (i.e., one day ahead) for the three types of the feed classified according to their weight. From figure 4.4b, the RMSE values are observed to be 0.33 m^3/day for heavy weight feed (i.e., feed weights greater than 20 kg), 0.46 m^3/day for medium feed (i.e., feed weights within the range 10-20 kg) and 0.39 m^3/day for light weight feed (i.e., feed weight less than 10kg). Considering the size of the dataset, these RMSE values obtained are relatively low with the heavy weight feed having the least range of errors compared to the other two feeds. Also, the coefficient of variance (CV), for the three types of feed showed that the heavy weight feed was the least compared to the other two types of feed. The results of the RMSE and CV indicate that the biogas estimation model is more sensitive to the feed with heavy weight compared to the feed with lower weights. However, the three RMSE and CV values obtained are quite low thereby indicating that the variation between the measured and predicted biogas is quite minimal. This thereby implies that the efficiency of the model developed is relatively high hence it is reliable to be used as a surrogate model for estimation of biogas generation in the micro-AD plant. Furthermore, the coefficient of variance (CV) in this research study was 13%. This implies a low spread of data values. This is most preferred as it suggests that the data values are quite close to the mean. However, the three RMSE values obtained which are relatively low indicate that the efficiency of the model developed is relatively high. Hence, it is reliable to be used as a surrogate model for estimation of biogas generation in the micro-AD plant.

Finally, an average relative RMSE of 2% was observed in Figure 4.4b. This average relative RMSE value obtained is very low, compared to other similar research studies, which were previously reported to be within a range of 5-10% by Wang *et al.,* (2020) and Pei *et al.,* (2022), 8.9%

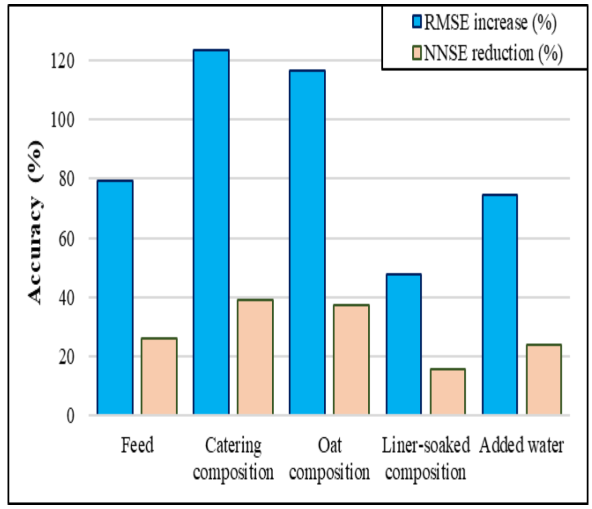
by Long *et al.* (2021), and 4.3% by Tufaner and Demirci (2020). This RMSE value obtained confirms the effectiveness of the developed model in predicting the biogas generated from the AD plant. It also indicates that the developed RNN/NARX model is robust enough to be used for relatively high fluctuated biogas generation. These research studies could successfully track the fluctuation by demonstrating more than 0.9 for coefficient of regression (R^2) or NSE (Tufaner and Demirci, 2020; Park *et al.,* 2021; Pei *et al.,* 2022), whereas the NNSE value obtained was 0.84 in this research study. Although the NNSE value obtained can be quite acceptable for this model but the high range of measured biogas (3.26±1.21 reported in Table 4.2) may be considered as the lack of perfect ability of the model to track biogas especially with the sudden rise or drop in biogas volume as shown in Figure 4.4a).



(a) **(b)**



(c)



(d)

Figure 4.4 (a) scatter plot of predicted biogas vs corresponding measurements for 1 day ahead, (b) comparison of observations with estimations, (c) analysis of the percentage of dataset used for model development, (d) Impact of feed compositions in the pre-feed tank on the biogas generation.

Figures 4.4c and 4.4d presents the results of the further analyses carried out on the developed RNN model where the results of both the uncertainty and sensitivity analysis are presented. As stated in the methodology, these analyses were significant to further confirm the effectiveness of the developed RNN/NARX model. The uncertainty analysis presented in Figure 4.4c implies that despite the minor higher resistance of NNSE in comparison to RMSE, the prediction accuracy of the RNN/NARX model, decreased with a corresponding decrease in the dataset.

While this finding was expected mainly because of the limited dataset (only 310 days) used for all steps of training, validation and testing the developed RNN model is still highly dependent on the volume of the dataset which has no unform pattern and correlation. This finding confirms the relatively lower coefficient of cross-correlation analysis input variables and the generated biogas, as illustrated in Figure 4.2. Figure 4.4d shows the results of the sensitivity analysis conducted by removing one decision variable and running model for one step ahead. From the sensitivity analysis presented in figure 4.4d, the catering composition was observed to have the highest level of accuracy compared to the other input variables. This shows that the developed RNN model is heavily dependent on the catering composition. The oats composition also demonstrated to have a very high level of accuracy and is the next to the catering composition. This indicates that both the oat and catering composition have the greatest significant impact on the biogas generation in the micro-AD plant compared to the other input variables.

The result of the sensitivity analysis also confirms the accuracy of the developed model as the analyses made by the operator of the micro-AD plant indicated that the catering composition demonstrated to have the highest composition of food waste comprising of 52% of the total food waste which had a major impact of the volume of biogas generated from the micro-AD plant (Walker *et al*.,2017). The analysis carried out by the operator also revealed that oats had a high composition of food waste compared to the other variables as it was next to catering comprising of 17% of the total food waste thus, the oats composition also had a significant impact on the biogas volume (Walker *et al.,*2017).

4.6 Periodic test and optimal results for the RNN/NARX model

SFLA optimisation method was used to specify the optimal calendar distribution and daily weights for each variable to maximise biogas generation. The optimisation method had 18 decision variables including four variables for feed added to the main digester at days t-3 to t, six variables for the water added to pre-feed tank at days t-5 to t, two variables for each of the three composition types (catering, oat, and liner) at days t-1 and day t-2, and biogas generation at day t and t-1.

The objective value is to maximise the biogas generation at day t+1. From table 4.3, the optimum values (i.e., optimum condition) for each of the aforementioned input variables were presented. The decision variables in the table are shown as predictors and the objective value is the estimation of biogas generation on the following day. As observed in the table, the estimation of maximum biogas generation was on day t+1 which is 4.52 m^3/day based on the 18 optimal decision variables at the preceding days between day t and t-3. This is the maximum possible volume of biogas that can be generated from this micro-AD plant based on the optimal values of decision variables. The analysis of the optimal decision variables shows that the entire feed for a cycle of four days can only be added on the last day (i.e. 80kg on day t), 60kg of catering is added to the pre-feed tank in two days, 55kg for day t-1 and 5kg for day t; 20kg of oat is added on day t-1; liner-soaked is not added and 15kg of water was added only on day t-1. To estimate the volume of biogas generated on the following day i.e., day t+2 based on the decision variables given in table 4.3 for the last 4 days i.e., between day t-2 and day t+1, the estimated biogas generation was observed to be 4.23 m^3/day. This decrease in the volume of biogas generated is due to the fact that the daily distribution of the input variable especially feed, catering and water is different from the optimal values obtained above. Similarly, the estimation of biogas generation in the following days (i.e., day t+3 and t+4) was observed to decrease further. On the other hand, if the same amounts of feed, catering and water are added every 4 days, the estimated biogas generation is repeated every 4 days. In other words, the estimation of the biogas values obtained after day 4 is observed to be similar to the predictor of the biogas values as the value t+3 is similar to the value t-1 and value t is similar to value t+4 accordingly. This indicates that the volume of biogas generated after every four days can be repeated with the same proportion of input parameters.

Another observation made in table 4.3 was the absence of the liner input variable in the generation of biogas volume. This shows that the liner variable had no significant impact on the maximum volume of biogas generated. This observation confirms the analysis made by the operator of the micro-AD plant where the liner composition was observed to be insignificant as little amount of liner was added to the digester thereby indicating that liner composition had no significant impact in improving the volume of biogas generated from the micro-AD plant. Also, it was observed from the table that no feed was made in the first two days of the operation (i.e., t-3 and t-2). Through this means, the micro-AD plant can be operated by local communities with minimum labour (i.e., most of the feeding is arranged for one day every -four days) to achieve the maximum efficiency of biogas generation.

The impact of this tends to address the problem of high operational costs associated with AD plants which is one of the economic challenges affecting the implementation of AD technology as a waste management option globally. This explains why cheap and unsustainable waste management practices such as landfills are still being utilised in many regions across the globe. Thus, the optimisation of the developed RNN model will lead to the optimisation of the total weight of feedstock processed to obtain maximum volume of biogas while reducing the operational cost of the micro-AD plant. Also, the efficacy of this tested approach demonstrates great potentials in making the micro-AD plant a more accessible and familiar option compared to the conventional AD plant as more people/investors will be encouraged to establish the micro-AD plant as an organic waste management technique in municipal areas. The effect of this has great tendencies in contributing towards the further reduction of transportation cost which is a major economic challenge associated with the utilisation of conventional AD plants for organic waste management (Walker *et al.,*2017). This implies that the developed AI and optimised model (RNN-SFLA) has the potential to address some of the economic challenges hindering the implementation of AD. It also demonstrates the ability of the developed RNN-SFLA model to improve organic waste management using AD technology thereby contributing towards the implementation of circular bioeconomy as the integration of RNN-SFLA into the micro-AD plant will help in encouraging more investors towards the establishment of more micro-AD plants.

The establishment of more micro-AD plants will bring development in rural communities across the globe through the promotion of small -scale industries in rural communities using clean energy generated from AD for various purposes. Also, the establishment of small– scale industries will help to create employment opportunities for people within the communities, thereby helping to reduce poverty within the area.

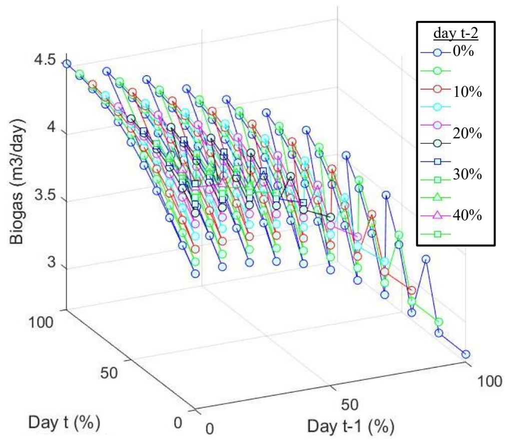
In addition, more organic wastes will be diverted from landfills through the establishment of more micro-AD plants for treating and managing organic wastes. This will help to minimise soil and water pollution which occur from the leachate of underground water pollution as well as potential run-offs due to the leaching of organic, inorganic, and various other substances of concern (SoC) contained in the waste. It will also help to minimise air pollution which occurs due to the suspension of particles. The diversion of more organic waste from landfills will also help to minimise odor pollution from the deposition of municipal solid waste (MSW). This will help to reduce health impacts which may occur through the pollution of the underground water and the emission of gases, leading to carcinogenic and non-carcinogenic effects of the exposed population living in their vicinity.

Table 4.3. Optimum condition for the operation of the micro-AD plant for maximum biogas generation

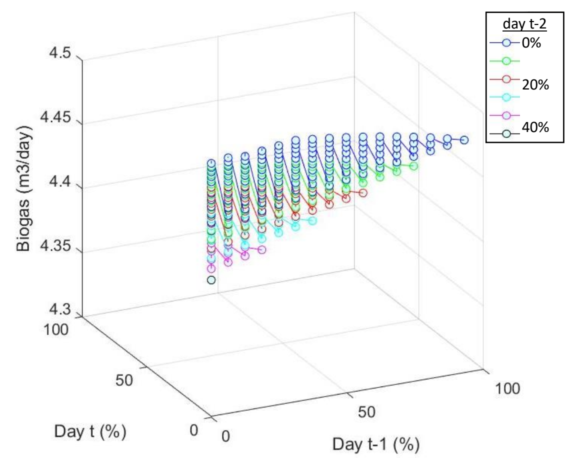
Parameter	**Days**							
	Predictors (input data)				**Predictions**			
	t-3	**t-2**	**t-1**	**t**	**t+1**	**t+2**	**t+3**	**t+4**
Feed	0	0	0	80				
Biogas	-	-	4.11	4.08	4.52	4.23	4.11	4.08
Catering	-	-	55	5				
Oat	-	-	20	0				
Liner	-	-	0	0				
Water	0	0	15	0				

Figure 4.5 shows the further sensitivity analysis carried out for biogas generation at day t+1 based on the percentage of variables over the past three days i.e., days t-2, t-1, and t for all the decision variables. This was to further evaluate the impact of the different input variables on the generated biogas. Figure 4.5a, shows the impact of the "feed to the main digester" on the volume of biogas generated was observed where different percentages of the feed data for days t and t-1 are shown in the horizontal axes and feed data for days t-2 are shown as graphs with an interval of 10%. As can be seen, the maximum volume of biogas generated (4.52 m^3/day) can only occur when AD is fed only on the last day (day t). In addition, every redistribution of feeding shows a decrease in

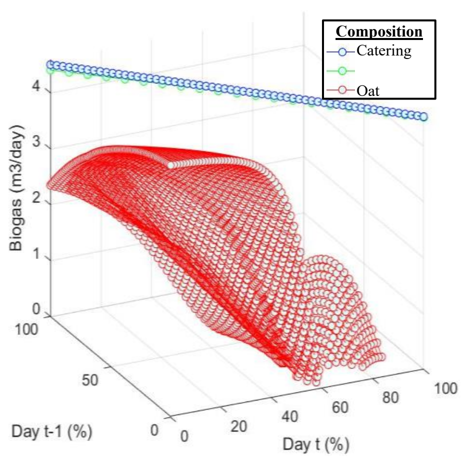
the biogas generation which can be translated as relative sensitivity of the model to the daily distribution of feed. For example, when feeding the AD plant in day t-1 instead of day t, the volume of biogas generated was observed to decrease from 4.5m^3/day to around 2.5 m^3/day (see blue circle in the left-up and down-right). The results also show that the model is highly sensitive to the amount of feed in last two days (i.e., day t and t-1) and the feed ratio for the other day, i.e., day t-2 that is unimportant (See blue circles vary more than the other lines thereby indicating that the model is sensitive to day t and day t-1).



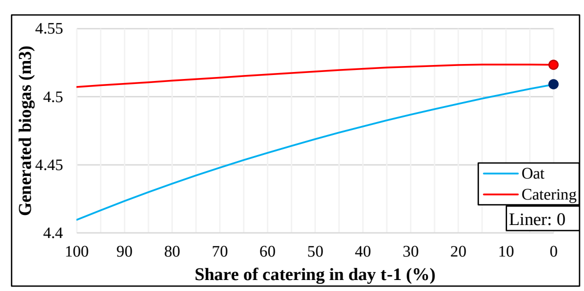
(a)



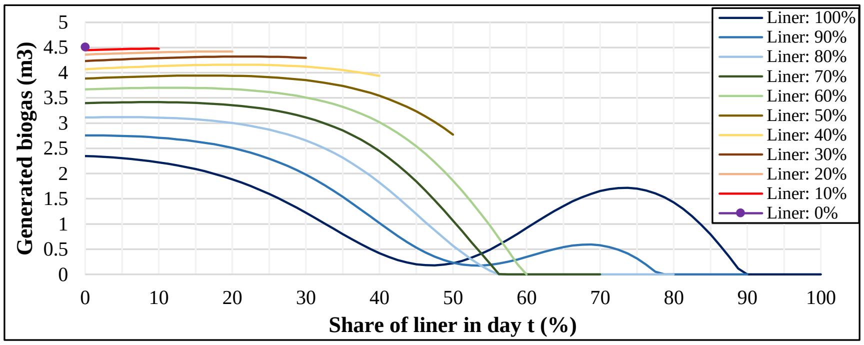
(b)



(c)



(d)



(e)

Figure 4.5. Biogas generation for the day t-2 distribution of the different decision variables: (a) feed, (b) water, (c) composition; and the impact of the distribution of the optimal values of the pre-feed composition variables on biogas generation for: (d) catering, oat (e) liner

Figure 4.5b shows the impact of the water added on the volume of biogas generation which was analysed at 20% intervals of the percentages. Compared to the feed distribution in Fig 4.5a, the daily distribution of added water is relatively unimportant for the model as the volume of biogas generated slightly increased by approximately 5% as it moved from 4.3 to 4.5 m^3/day.

Figure 4.5c presents the impact of the three different composition variables on the volume of biogas generation. It can be seen that the catering composition added to the pre-feed tank results in the maximum volume of biogas amongst other variables. This implies that the catering composition has a higher influence on biogas generation compared to the other two composition variables. This also confirms the analysis made by the operator of the micro-AD plant. Following this, the oat composition also generated a high volume of biogas as shown in Figure 4.5c. This also indicates that the oat composition has a strong influence on the volume of biogas generation while the liner composition had no significant influence of the volume of biogas generation. This is also in line with the sensitivity analysis presented in Figure 4.5d where both the catering and oat compositions had high volumes of biogas while the liner composition generated no significant biogas volume. In figure 4.5d, that the model shows a higher level of sensitivity to the addition of liner than the daily distribution of composition. For example, while daily distribution of oats and catering has no significant impact on the generated biogas, it is highly sensitive to the amount of added liner. When applying optimal operation strategy, it is crucial to understand the significance of the distribution of composition variables added to the pre-feed tank over the cyclic period.

Hence, the impact of distribution of the optimal amount of the pre-feed composition variables on biogas generation is further analysed in figure 4.5d-e for three individual variables i.e., catering, oats, and liner separately. More specifically, Figure 4.5d considers distribution of optimal value of the catering variable at day t where other variables are fixed here as 20kg for oat zero for liner and optimum condition of accumulative rate of catering is 60kg (i.e., 55kg for day t-1 and 5kg for day t). As can be seen in Figure 4.5d, the biogas generation decreased for other distribution rates down to 4.515 m^3. This indicates that while the AD plant is highly dependent on catering (as shown in the sensitivity figures), its distribution between the day t and t-1 has no significant impact on biogas generation. Although there was a drop in the volume of biogas generation as the share of oat on day t increased compared to the optimum value, the drop in biogas volume has no significant impact on biogas generation. In Figure 4.5e, the share of liner in day t (%) is observed where each line corresponds to a percentage of liner in total waste. The horizontal axis shows how this percentage is distributed between day t and day t-1. Hence, for 20% of the liner, the available data is for 0-5-10-15 and 20%. From Figure 4.5e, it can be observed that the increasing liner causes a decrease in the volume of biogas as the lines move closer to 100% after line 50%.

Finally, the strategy for the optimal generation of biogas was compared with best feeding events and entire test period (47 days). As it can be seen in figure 4.6a, the biogas generation in all three best identified feeding events, is relatively similar and a uniform increase with a maximum weekly volume of 26.14 m^3. However, the maximum volume of biogas generation for the optimised operation is 29.97 m^3 i.e., an improvement rate of approximately 15%. In figure 4.6b, the average volume of biogas generated for the first 40 days were observed to be 3.26 m^3/day. It was also observed that the average volume of biogas generated for the 47-day test period was 4.34m^3/day. The average volume of biogas generated for the entire 47-day test period indicates that the RNN-

SFLA model has the potential to improve the performance of the micro-AD plant as the average volume of biogas generated from the micro-AD plant by the operator was observed to be $3.15m^3$/day.

However, the generated biogas experienced a decrease from days 40-46 as the average volume of biogas generation between days 41-46 was observed to be 1.48 m^3 /day. Similarly, the generated biogas volume for the entire test period for the measured event shows an increase up until day 40 when it experienced a slight decrease in the generated biogas volume. It then increased the next day with a maximum volume of 139.51 m^3.

On the other hand, the generated biogas volume in the optimised operation can uniformly increase with a steeper slope and achieve up to a maximum of 199.46 m^3 i.e., a significant improvement of approximately 43% for biogas generation compared to business-as-usual. The proposed model shows an outstanding performance in both short- and long-term operation, i.e., 7 days and 47 days in which longer period results in more volume of biogas enhancement. The ability of the developed RNN-SFLA model to improve biogas generation by 15% and 43% over a period of 7 and 47 days respectively indicates the effectiveness of the RNN-SFLA model in improving biogas generation from the micro-AD plant. It also demonstrates the benefit of developing an optimised strategy like SFLA for the operation of the micro-AD plant having the potential to improve the overall performance and productivity of the micro-AD plant thereby resulting in a considerable increase in the amount of biogas volume generated. This helps to overcome the critics made on the technical limitations of AD such as low yield and slow operation which has often discouraged the further application of the technology for the management of organic wastes. The application of SFLA as an optimisation tool is particularly important for future research works in AD as the application of SFLA as an optimisation tool in AD operations is a novel approach which to the best of the author's knowledge, has never been applied within the context of AD.

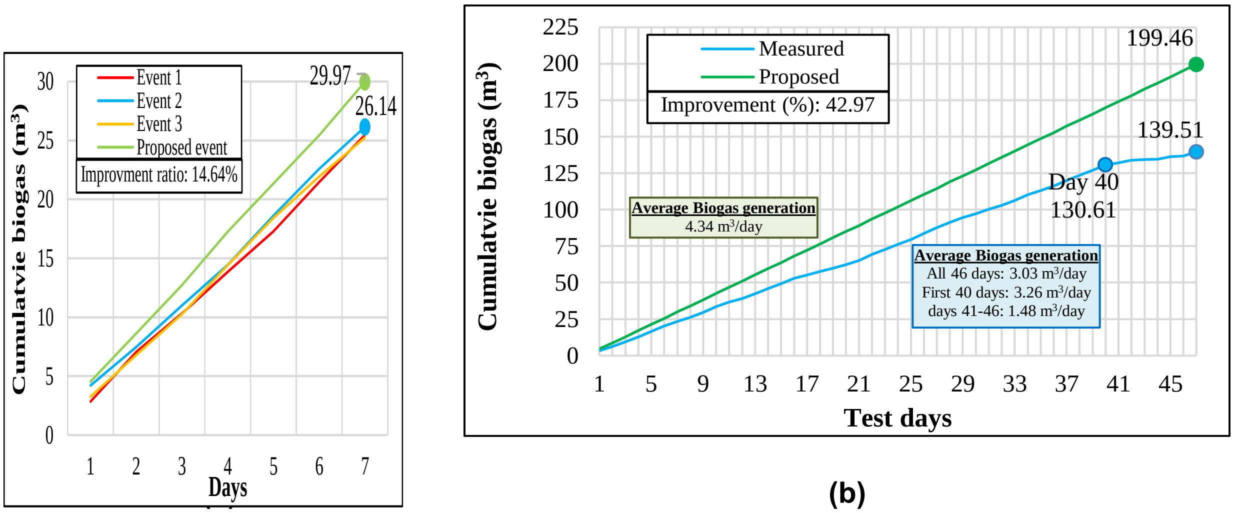


Figure 4.6. Comparison between the best feeding events and the proposed (optimised) operation for operation in (a) 7 days and (b) 47 days

4.7 Significance of the developed RNN Model

The RNN model was developed to improve the performance of a micro-AD plant in producing maximum volume of biogas. The development of the RNN model using a real micro-AD plant was a significant advancement in the field of AI applications in AD systems globally. This is because previously developed models mainly used simple ML or ANN which were mostly applied on laboratory scales thereby limiting their application in the context of industrial setting. The use of simple ML or ANN on laboratory scales has also limited their widespread deployment. The developed RNN model demonstrated promising potential in the effective prediction of biogas. This was indicated by the performance indicators used in the research study (RMSE and NNSE). It also demonstrated to have great potential in improving the performance of the micro-AD plant through the production of maximum volume of biogas. This was indicated over a 7- and 47-day period as presented in figure 4.6.

The ability of the developed RNN model to predict biogas accurately from a micro-AD plant is highly relevant for various key reasons. It gives room for effective energy planning and management by providing decision makers with the means to ascertain the potential energy output.

This information is vital for assessing the feasibility and profitability of implementing these systems on a larger scale (Wang *et al.,* 2020). Secondly, the optimisation of biogas produced from the micro-AD plant using SFLA gives operators an insight into the mechanisms behind optimal biogas production. This information is vital as it enables the operators to adjust key operational variables for the purpose of ensuring the efficient utilisation of feedstock. It also helps to reduce the operational cost, improve the efficiency of the process, and minimise the risk of system failures while increasing biogas produced from the AD system (Khan *et al.,* 2023).

Through the application of the RNN model in the AD system for the above-mentioned reasons, some of the technical and economic challenges associated with the application of AD as a waste management option can be tackled. This will assist in making the AD system a more suitable waste management option within the global context, especially in regions around the world where the application of AD technology is yet to be implemented due to some of these challenges. This approach will contribute towards the implementation of circular bioeconomy. It will also contribute towards realizing the potential of AD technology in meeting up with the ever-increasing energy demands of the people globally.

4.8 Results of the Feature Extraction and Selection Analysis

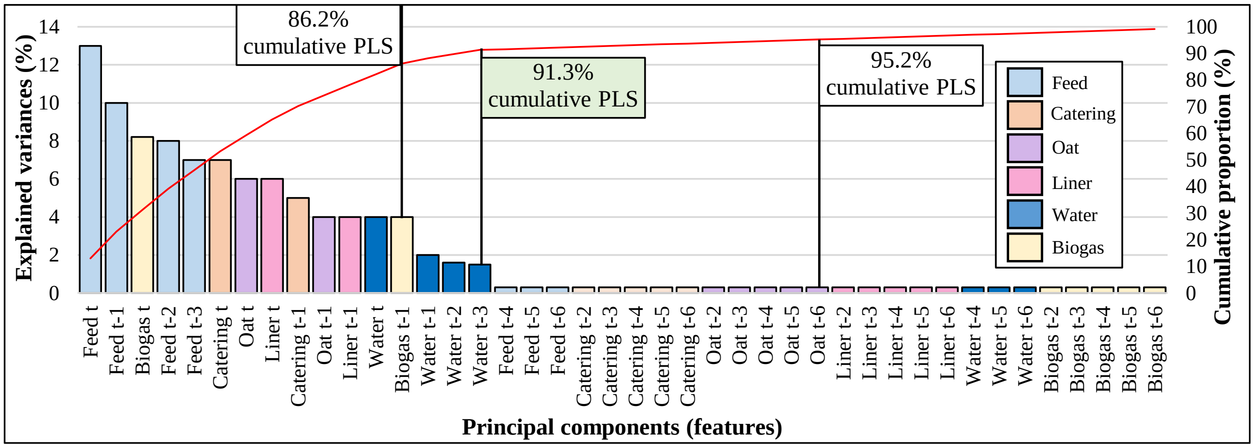
Figures 4.7a and b presents the results of the PCA, PLS and sequential sensitivity analysis (SSA) conducted for all the group features outlined in Table 3.1 in the methodology section From the results of the PCA and PLS analysis presented in figure 4.7a, it was observed that out of the forty-two (42) total time-series features, fifteen (15) features account for over 90% of the cumulative explained variances (91.3% in Figure 4.7a). In addition, the feeding of the last four days (t to t-3) contributed significantly to this group of features. This indicates its high impact on the modelling process.

This observation was further corroborated by the sequential sensitivity analysis shown in Figure 4.7b where the feed feature was observed to have the highest impact on biogas volume compared to the other features. This indicates the significance of the feed feature in volume of biogas generated from the AD plant.

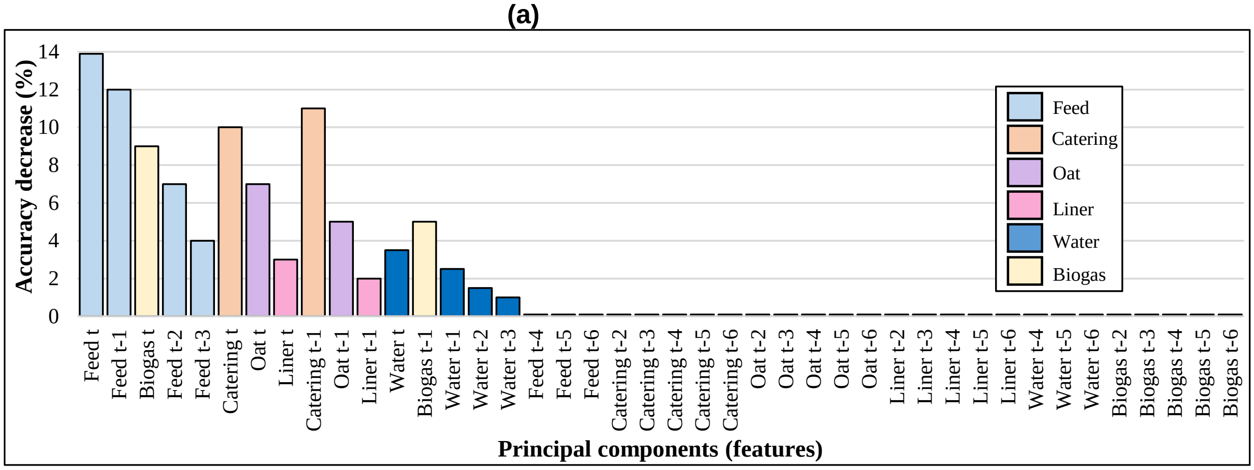
In addition, the biogas levels at time t and t-1 were observed to exhibit a substantial impact on the modelling outcomes. This suggested that some part of biogas production may be influenced by the feeding activities of the current day and the day before for prediction of yielded biogas for next day. Moreover, the analysis demonstrates that the waste composition during the last two days significantly affects both the PLS analysis and the accuracy of WLDM modelling. Finally, the analysis identifies the last four days of added water to the pre-digester as the most influential factor. The overall importance of the features obtained from the PLS analysis appears to align well with the results of the sequential sensitivity analysis, specifically when considering the criterion of cumulative PLS over 90%. This implies that the features selected based on their cumulative PLS values above 90% indeed have a significant impact on the accuracy of the model.

However, it is important to recognise that changing the criterion for cumulative PLS may lead to different results. For instance, if the criterion is set to cumulative PLS over 95%, additional features with negligible impact on accuracy may be included in the analysis, thereby making the model less efficient in practice, Comparing figure 4.7a with 4.7b for more reference. On the other hand, if the cumulative PLS criterion is lowered to 85%, relevant features with a noticeable impact on accuracy, such as the three features related to added water, might be excluded, leading to potential loss of predictive power. Therefore, relying solely on the PLS analysis, which is commonly utilised in series of research works, might not always yield the most accurate or appropriate results.

Thus, the incorporation of the sequential sensitivity analysis mentioned earlier, is vital in the feature selection process. Such an approach allows for a more robust comprehension of the features' actual influence on the accuracy of the model. It also helps in making informed decisions as regards their inclusion or exclusion. Through this means, the reliability or suitability of the final model for practical applications is increased. Finally, based on the feature analysis, among all the extracted time-series features, the last four days of digester feeding, added water to the pre-digester, the biogas generation data for two days, as well as information regarding oats, liner and catering added in last two days were selected to develop the final WLDM models and the proposed ensemble model thereafter.



(a)



(b)

Figure 4.7. Feature analysis of input data: (a) PLS analysis of extracted features, (b) average accuracy decrease of all WLDM obtained by sequential sensitivity analysis.

4.9 Performance of the WLDM models

The results of performance of the six different WLDM models are presented in Figure 4.8, based on three KPIs: TPR, TNR, and ACC. The confusion matrix of these models can be found in Figure

A1 in the Appendix. For the TPR of the low class (Figure 4.8a), the DA, DT, and NB models were observed to demonstrate an acceptable rate in the prediction of low class of biogas. This indicates their ability to correctly identify instances in this class however, the DT model had a higher performance than the DA and NB models. The superior performance of the DT model over the DA and NB models further confirms the effectiveness of the DT model to predict biogas accurately as previously revealed in different research studies conducted by different researchers. For example, in a study by Wang *et al* (2021a), DT was reported to be an effective ML tool for predicting biogas produced from AD. The research study by Wang *et al* (2021a) also revealed that DT performed better than the NN models due to its high dimensional adaptive feature learning capability. Another research study conducted by both Cheon *et al.,*2022 and Park *et al.*,2021 respectively revealed the ability of DT models to further improve the predictive accuracy of AD variables. Similarly, recent research studies conducted by Gupta *et al* (2023) also confirmed the effectiveness of DT in predicting biogas and other AD output variables.

Further research studies by Gupta *et al* (2023) also reported that an ensemble DT model had successfully been used to predict the transient VFA accumulation in AD reactors which is highly detrimental to biogas production. In the case of the medium class (Figure 4.8b), GPR, DA and DT demonstrated acceptable rates of predicting the medium class of biogas. However, the GPR model was observed to have a superior performance than the other WLDM models (DA and DT) in predicting biogas accurately. This confirms the effectiveness of the GPR model in predicting the accuracy of biogas. Though limited research studies conducted previously have applied the GBR model in predicting the accuracy of biogas (Gupta *et al.,*2023).

However, research studies by Trucchia & Frunzo (2021) reported that GPR-based surrogate modelling has been used for the quantification of uncertainty and global sensitivity analysis of modified ADM1 which was observed to be capable of predicting both CH_4 production and VFA accumulation in AD processes. Also, the GPR based surrogate model has been revealed to rank different model parameters based on their relative impact (Gupta *et al.,*2023).

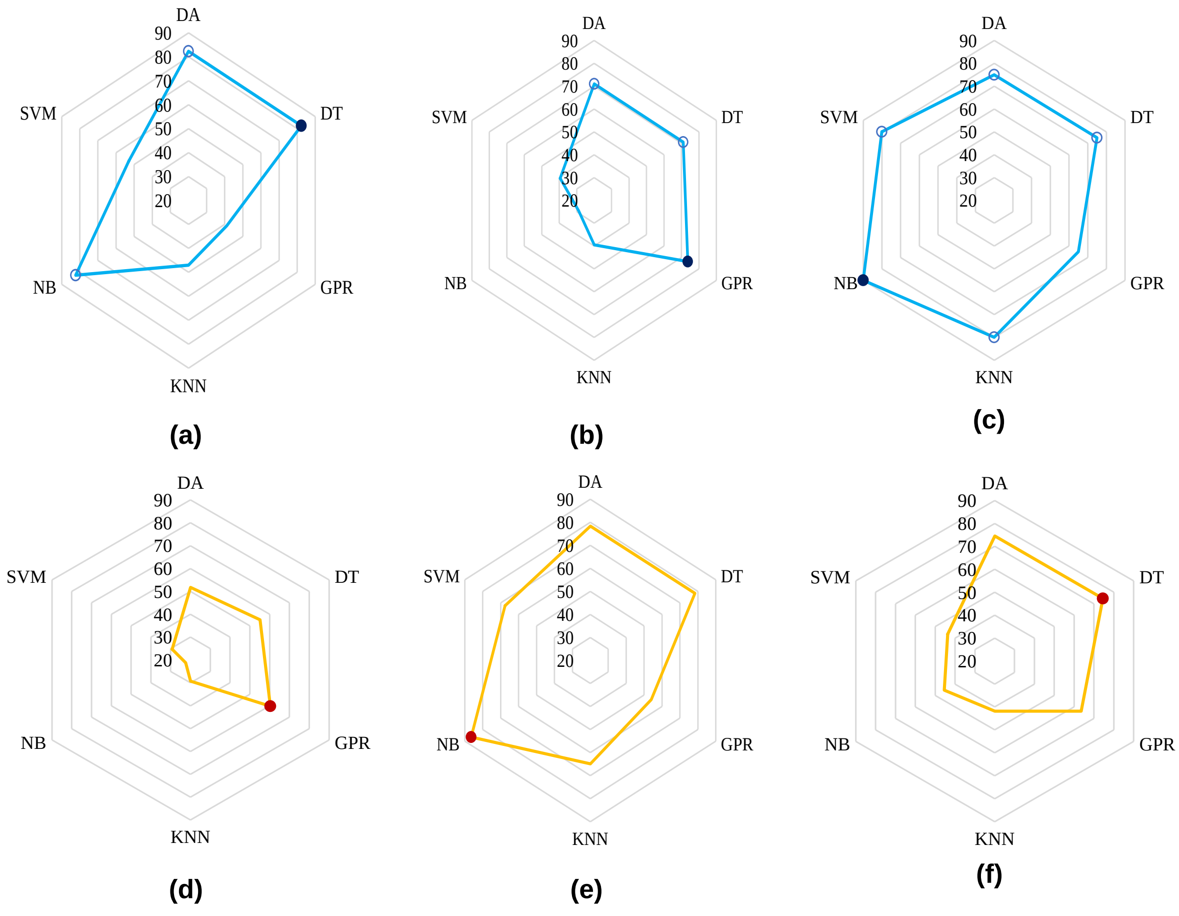
The performance of GPR-based surrogate modelling with FNN and the polynomial chaos expansion (PCE) (a method of expressing a random variable as a polynomial function of other random variables) were compared for a WWTP with an AD unit (i.e., BSM2) (Al *et al*., 2019). The results obtained showed that GPR-based global sensitivity analysis had a better performance the FNN and PCE models in terms of training time. This can be attributed to the fact that they required lower number of datapoints during training.

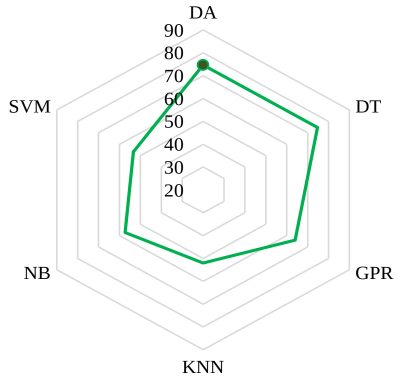
Furthermore, for the high class of biogas, (Figure 4.8c), all the WLDM models, except for the GPR model were observed to exhibit excellent performance in recognising situations with high yielded biogas. However, the NB model was observed to perform better than the other WLDM models (i.e., SVM, KNN, DT and DA). The performance of these models confirms their effectiveness in the prediction of biogas as models such as NB, SVM, KNN, DT and DA have been revealed in previous research works. For example, De Clercq *et al* (2019) employed different machine learning models (SVM, logistic regression (LR), RF, XGBoost, and KNN) in the enhancement of biogas produced from industrial facilities. This was achieved by designing a graphical user interface to the machine learning models capable of predicting biogas output given a set of waste inputs. The study by De Clercq *et al* (2019) revealed that KNN obtained the highest prediction accuracy compared to the other machine learning models. Recently, Yildirm & Ozkaya (2023) employed five different machine learning (ML) algorithms (RF, ANN, KNN, SVR, and XGBoost) in describing and predicting the correlation between the operational parameters and the quantity of generated biogas collected from a real-scale anaerobic digestion plant. The research study by Yildirim & Ozkaya (2023) revealed that both SVR and KNN had high total biogas prediction accuracies of 0.8655 and 0.8326 over time respectively. Similarly, Wang *et al* (2020) applied KNN, SVM, GLMNET and RF in the prediction of methane yield from biogas where KNN had the highest prediction accuracy. Alejo *et al* (2018) developed a C-SVM model to predict the effluent composition of the two-stage AD process with poultry manure as a feedstock. The accuracy of C-SVM was compared with other predictive models based on FNN and stoichiometric analytical methods, where C-SVM showed superior prediction accuracy. Long *et al* (2021) investigated the feasibility of six ML algorithms (SVM, GLMNET, RF, KNN, NNET and extreme gradient boosting (XGBOOST) in improving biogas production through the prediction of methane yield where SVM was reported to have the second highest degree of accuracy after RF. Similarly, Onu *et al* (2023) confirmed the effectiveness of SVM in the accurate prediction of biogas and other output variables from AD.

Though, there has not been many applications of DA in the prediction of biogas from AD however, DA was applied together with UV/vis spectroscopic probes in the prediction of organic acid from a biogas plant where it was reported to have an accuracy of more than 87% (Wolf *et al.,*2010). Another research study conducted by Molina *et al* (2009) revealed that factorial discriminant analysis (FDA) was successfully applied together with phenomenological analysis to characterize steady states and dynamic response analysis against disturbances which occur during the systematization of indicators for two types of wastewaters. Brambilla *et al* (2012) revealed that linear discriminant analysis (LDA) was successfully used in multivariate statistical data processing in conjunction with principal component analysis (PCA) to monitor biogas and methane production from an AD plant under different operational parameters .Limited research studies have shown the application of the NB model in AD systems however, NB has been revealed as an effective classifier having been successfully applied in previous research works for different classification purposes (Wickramasinghe &., Kalutarage,2021).

On the other hand, in figure 4.8d, the GPR model was the only WLDM model observed to have TNR for low class of biogas. In figure 4.8e, only the NB model demonstrated to have TNR for medium class of biogas. Figure 4.8f indicated that only the DT model was observed to have TNR for high class of biogas. When comparing figures 4.8d-f with figures 4.8a-c, it becomes evident that although the models excel in detecting the high-class station, they also display many instances of underestimation and overestimation in other situations, as evident from the low TNR in the low class and high class, which are only relatively compensated in the medium class (Figure 4.8e).

These limitations result in low overall accuracy for all the WLDM models in the prediction of biogas. This includes the best-performing one, DA, as shown in Figure 4.8g, where it achieves an accuracy close to 80%. These findings suggest that while the selected models demonstrate to be proficient in certain areas, they still suffer from some shortcomings which hinder their overall accuracy. Consequently, the proposed ensemble model combined the strengths of the superior WLDM models in each class to improve the prediction performance in a broader range of scenarios.





(g)

○ Selected models ● Best model

Figure 4.8: Performance of developed WLDM models: (a) TPR of low class, (b) TPR of medium class, (c) TPR of high class, (d) TNR of low class, (e) TNR of medium class, (f) TNR of high class, (g) ACC rate

Figure 4.9 presents further illustrations on the performance of each of the WLDM models in the prediction of the three different classes of biogas. This was illustrated in the results of confusion matrix for each of the WLDM models. From figure 4.9a, it was observed that the DA model had an 82% correct estimation of the low class of biogas. It was also observed that it had 18% over estimation for the low class of biogas. For the medium class of biogas, DA had a correct estimation of 71%. It also had over estimation of 18% and under estimation of 11%. In the case of the high class of biogas, DA had a correct estimation of 75% and an under estimation of 15%. The performance of DA further confirms its effectiveness in the accurate prediction of the three different classes of biogas.

In figure 4.9b, the DT model was observed to have a slightly similar performance with the DA model in the effective prediction of the different classes of biogas. It had a correct estimation of 82% and an over estimation of 18% for the low class of biogas like DA model. In the case of medium class of biogas, the DT model obtained a 71% correct estimation of biogas like the DA model. It also under estimation of 13% and an over estimation of 16%. The performance of the DT model further confirms the effectiveness of the DT model in predicting the different classes of biogas as reported previously.

In figure 4.9c, the GPR model had a correct estimation of 41% for low class of biogas and an overestimation of 59%. For the medium class of biogas, GPR demonstrated to have 74% correct estimation, 8% under estimation and 18% over estimation. In the case of high class of biogas, the GPR model gave 65% correct estimation and 35% under estimation. The performance of the GPR model in the three different scenarios of biogas indicates that it can provide a better and more effective prediction of the medium and high classes of biogas compared to the low class of biogas.

Figure 4.9d showed that the KNN model had a correct estimation of 47% and an over estimation of 53% for low class of biogas. For the medium class of biogas, it was observed to have a correct estimation of 39%, an under estimation of 8% and an over estimation of 53%. In the case of the high class of biogas, the KNN model had a correct estimation of 80%, an under estimation of 10% and an over estimation of 10%. The performance of the KNN model in the three different scenarios implies that it can effectively predict the high class of biogas with minimal under estimations and over estimations respectively.

Figure 4.9e demonstrated the performance of the NB model in the prediction of the three different classes of biogas. In the case of the low class of biogas, NB was observed to have correct estimation of 82%. It also had an under estimation of 18%. For the medium class, NB had a correct estimation of 29%, an under estimation of 21% and an over estimation of 50%. In the case of the high class of biogas, NB had a correct estimation of 90% and an under estimation of 10%. Similar to the KNN model, the NB model demonstrated a high level of effectiveness in the accurate prediction of high class of biogas compared to the low and medium classes of biogas.

In figure 4.9f, the SVM model had a correct estimation of 53% and an over estimation of 47% for the low class of biogas. For the medium class, it had a correct estimation of 39%, an under estimation of 8% and an over estimation of 53%. For the high class of biogas, SVM had a correct estimation of 80% and an under estimation of 20%. The performance of the SVM model also demonstrates its effectiveness in the prediction of high class of biogas compared to the low and medium classes of biogas. The performance of each of the WLDM models in the confusion matrix results presented further confirm the effectiveness of the DA model which supersedes other WLDM models. However, the overestimation and underestimation of the predictions made remains a challenge hindering the overall performance of the WLDM models.

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	14	3	0
Medium	4	27	7
High	2	3	15

(a)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	14	3	0
Medium	5	27	6
High	0	5	15

(b)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	7	10	0
Medium	3	28	7
High	0	7	13

(c)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	8	9	0
Medium	3	15	20
High	2	2	16

(d)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	14	1	2
Medium	8	11	19
High	0	2	18

(e)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	9	8	0
Medium	3	15	20
High	2	2	16

(f)

Over estimation | Correct estimation | Under estimation

Figure 4.9 Confusion matrix result of WLDM models: (a) DA, (b) DT, (c) GPR, (d) KNN, (e) NB, (f) SVM

Figure 4.10 also shows further confusion matrix results of the WLDM models comprising of the proposed model, the low based, medium based, high based, hard voting, soft voting, RF, subspace NB, XGBoost, gentle boost DA and RUS boost GPR. Aside from the proposed model, other models form the bagged, stacked and boost models respectively. Each of these models were compared with the proposed model to determine the effectiveness of the proposed model in improving the prediction accuracy of biogas.

From figure 4.10a, the proposed model was observed to have 90% correct estimation and 2% over estimation of the low class of biogas. It was also observed to have 88% correct estimation, 8% over estimation and 4% under estimation for the medium class of biogas. In the case of the high class of biogas, the proposed model obtained 93% correct estimation and 7% under estimation.

In figure 4.10b, the low based model had 85% correct estimation and 15% over estimation for the low class of biogas. For the medium class of biogas, it had 72% correct estimation, 8% under estimation and 20% over estimation. In the case of the high class of biogas, the low based model had 53% correct estimation and 47% under estimation. The results of performance of the low-based model in predicting the three different classes of biogas gives a strong indication of the proficiency of the low-based model in predicting low class of biogas compared to the medium and high class of biogas respectively.

Figure 4.10c showed that the medium based model had a correct estimation of 50% and an over estimation of 50% for the low class of biogas. For the medium class of biogas, it had a correct estimation of 84%, an under estimation of 4% and an over estimation of 12%. For the high class of biogas, the medium based model demonstrated to have a correct estimation of 47%, and an under estimation of 53%. The results obtained in figure 4.10c indicate that the model is most proficient in the accurate prediction of the medium class of biogas compared to the low and high class of biogas.

In figure 4.10d, the results obtained high based model were observed to be most proficient in the accurate prediction of the high class of biogas compared to the low and medium classes of biogas respectively. It was observed to have a correct estimation of 50% and an over estimation of 50% in the case of low class of biogas. For the medium class, the high based model had a correct estimation of 68%, an under estimation of 4% and an over estimation of 28% while it had a correct estimation of 87% and an under estimation of 13% for the case of high class of biogas.

Figure 4.10e showed that the hard voting model had a correct estimation of 30% and an over estimation of 70% for the low class of biogas. In the case of the medium class, it was observed to have a correct estimation of 76%, over estimation of 12% and under estimation of 12%. For the high class, it had a correct estimation of 43% and under estimation of 57%. The performance of the hard voting model presented in figure 4.10e implies that it is most proficient in giving more accurate predictions of the medium class of biogas compared to the low and high class of biogas.
In figure 4.10f, the soft voting model had a correct estimation of 45%, and an over estimation of 55% for the low class of biogas. For the medium class of biogas, it was observed to have a slightly better prediction accuracy as it demonstrated to have a correct estimation of 56% an under estimation of 28% and an over estimation of 16%. In the case of high class of biogas, the soft voting model obtained a much higher prediction accuracy than the low and medium class of biogas as it had a correct estimation of 80% and an under estimation of 20%.
Figure 4.10g showed that the RF model had correct estimation of 83% and an over estimation of 17% for the low class of biogas. For the medium class, it had a correct estimation of 72%, an under estimation of 8% and an over estimation of 20%. For the high class, the RF model had a correct estimation of 73% and an under estimation of 27%. The performance of the RF model gives a relatively strong indication of its effectiveness in the prediction of the three classes of biogas however, it tends to be more proficient in the accurate prediction of the low class of biogas than the medium and high class of biogas.
In figure 4.10h, subspace NB demonstrated to have a correct estimation of 65% and an over estimation of 35% for the case of low class of biogas. For the medium class, it was also observed to have a correct estimation of 64%, an under estimation of 16% and an over estimation of 20%. In the case of high class, subspace NB had a correct estimation of 87% and an under estimation of 13%. The performance of subspace NB implies that it is most effective in the accurate prediction of high class of biogas compared to the medium and low class of biogas.

Figure 4.10i showed that XGBoost had a correct estimation of 78% and an over estimation of 22% for the low class of biogas. For the medium class of biogas, XGBoost was observed to have an underestimation of 4%, a correct estimation of 76% and an overestimation of 20%. In the case of high class of biogas, it had a correct estimation of 77% and an under estimation of 23%. The performance of XGBoost indicates that it is proficient in the accuracy in the prediction of the three classes of biogas. However, it demonstrates a higher level of proficiency in the accurate prediction of the low class of biogas.

In figure 4.10j, gentle boost DA was observed to have a correct estimation of 77% and an over estimation of 23% for the case of low class of biogas. For the medium class of biogas, it had a correct estimation of 48%, an under estimation of 20% and an over estimation of 32%. For the case of high class of biogas, gentle boost DA had a correct estimation of 82% and an under estimation of 18%. The results presented in figure 4.10j indicate the high effectiveness of gentle boost DA in giving accurate predictions of the high class of biogas compared to the low and medium class of biogas.

Figure 4.10k showed the performance of RUS Boost GPR over the three different classes of biogas. For the case of the low class of biogas, RUS Boost GPR was observed to have a correct estimation of 55% and an over estimation of 45%. It had a correct estimation of 72%, an under estimation of 16% and an over estimation of 12% in the case of the medium class of biogas. In the case of the high class of biogas, RUS Boost GPR had a correct estimation of 83% and an under estimation of 17%. RUS Boost GPR demonstrated relatively high performance in the accurate prediction of the medium and high class of biogas compared to the low class of biogas whose performance was observed to be slightly above average. Generally, the proposed model was observed to have the best performance in the accurate prediction of the three classes of biogas compared to the other WLDM models. It was also observed to have minimal over estimation and under estimation compared to the other WLDM models. Though the occurrence of over estimation and underestimation remains a challenge, the performance of the proposed model further confirms the effectiveness of the proposed model in achieving the aims and objectives of this research study.

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	18	1	1
Medium	1	22	2
High	1	1	28

(a)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	17	2	1
Medium	2	18	5
High	3	11	16

(b)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	10	9	1
Medium	1	21	3
High	5	11	14

(c)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	10	9	1
Medium	1	17	7
High	1	3	26

(d)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	6	14	0
Medium	3	19	3
High	0	17	13

(e)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	9	9	2
Medium	7	14	4
High	1	5	24

(f)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	15	2	1
Medium	2	18	5
High	2	6	22

(g)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	13	6	1
Medium	4	16	5
High	2	2	26

(h)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	14	1	3
Medium	1	19	5
High	1	6	23

(i)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	14	2	2
Medium	5	12	8
High	3	3	27

(j)

Measured biogas class \ Prediction biogas class	Low	Medium	High
Low	11	7	2
Medium	4	18	3
High	1	4	25

(k)

Over estimation | Correct estimation | Under estimation

Figure 4.10. Confusion matrices result of WLDM models: (a) proposed, (b) low based, (c) medium based, (d) high based, (e) hard voting, (f) soft voting, (g) RF, (h) subspace NB, (i) XGBoost, (j) gentle boost DA, (k) RUS boost GPR.

4.10 Outcome on the Performance of the Developed Ensemble Model

The results of the performance of the proposed ensemble model and other benchmark models are presented in figure 4.11. From the results presented in figure 4.11, it was observed that the accuracy of the proposed ensemble model was observed to be 91%. Comparing the performance of the proposed ensemble model to the DA model in Figure 4.8g, it can be deduced that there was an 11% improvement in the prediction accuracy of biogas. (i.e., from 80% obtained in DA model in Figure 4.8g to 91%). This shows an outstanding performance of the proposed ensemble model in comparison to the other benchmark models with a remarkable 4.5% for each underestimation and over estimation while the other benchmark models had relatively higher underestimation and overestimation. This finding indicates that the proposed ensemble model could perfectly fill the knowledge gap by improving the accuracy of biogas predictions. The ability of the proposed ensemble model to improve the accuracy of biogas predictions implies that the ensemble model can improve the overall performance of the micro-AD plant in generating biogas. Also, it was observed that both hard and soft voting of all stacked models showed accuracies around 63% and 51% respectively. The performance of both hard and soft voting of all stacked models. Comparing the results of the hard and soft voting of all stacked models with the result of DA model presented in figure 4.8g, it can be deduced that using all the WLDM models to improve the prediction accuracy of biogas is a less meaningful approach as it cannot improve the prediction accuracy of biogas effectively. On the other hand, group stacking of the models based on their capability in specific class, for example low or high class shows better result, compared to the hard and soft voting of all stacked models especially for high-based model which could improve the prediction accuracy to almost 70%. However, the high overestimation of the benchmark stacked models is still challenging, especially for optimal operation in which higher rate of yielded biogas is the goal. The results reveal a clear superiority of the bagging and boosting models, particularly RF and XGBoost, over the benchmark stacking models. This finding aligns with previous research conducted on numerical problems (Xu *et al.,* 2021; Sonwai *et al*., 2023). However, it's important to note that despite this progress, the accuracy achieved, which remains below 80%, still falls short when compared to the performance of the proposed stacked model. To validate these outcomes, the research study examined the receiver operating characteristic (ROC) curves and their corresponding area under the ROC curve (AUC) for the top four performing models as presented in figures 4.11 b and c respectively.

Detailed comparisons between the performance of the developed ensemble models and the WLDM models can be depicted in Figure 4.8 b-g. The proposed ensemble model notably exceled by consistently maintaining an AUC above 0.94 across all classes. It particularly demonstrated an exceptional performance in the high class where the AUC was reported as 0.98 (Figure 4.9b). In contrast, the alternate benchmark models exhibited AUC values ranging from 0.74 to 0.8 in the high class, while their performance notably deteriorated in the medium class with AUC figures of 0.62 to 0.72. Moreover, it is worth noting that the optimal thresholds for the proposed model remained relatively consistent along the x-axis, within the range of (0.8-1, 0.2-0.1), while for the other models, these thresholds shifted towards higher values.

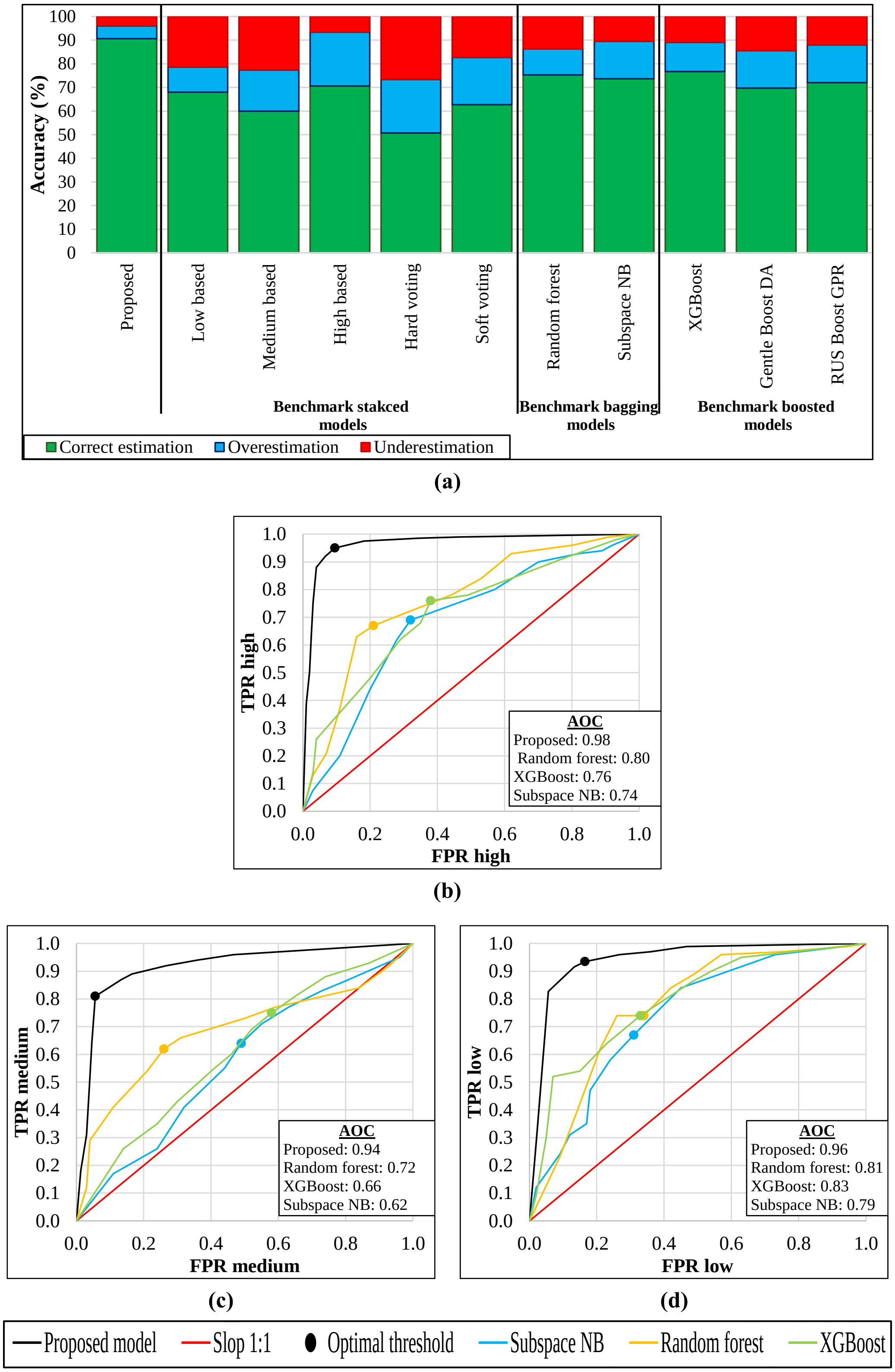


Figure 4.11 Performance of different benchmark ensemble models in comparison to proposed model.

Table 4.4 presents further details on the performance of the developed ensemble models for the benchmark stacked models, benchmark bagged models and benchmark boosted models in giving both TPR and TNR for the three classes of biogas (i.e., low, medium, and high) respectively. The table presented below, also showed the degrees of accuracy for all the benchmark models. From the table presented below, the proposed model was observed to have the highest TPR and TNR for the three different classes of biogas compared to the other benchmark models presented below. The effect of this gave rise to the high degree of accuracy obtained. The result of the proposed model presented in table 4.4, provides a more comprehensive justification on the accuracy value obtained (i.e., 91%).

Another observation made in the performance of the developed ensemble models is the degree of accuracy obtained for the benchmark stacked, benchmark bagging and benchmark boosted models respectively. From the performance of the benchmark models, it was observed that both bagging and boosted models had better performances than the stacked models as they had relatively higher degrees of accuracy compared to the stacked models.

The bagged and boosted models were observed to have accuracies within the range of 70%-77% while the stacked models had accuracies within the range of 51%-71% in the prediction of biogas. This indicates that the performance of both the bagged and boosted models supersedes the performance of the stacked model in predicting biogas accurately. However, despite the performance of both the bagged and boosted models, it still had a lower performance than the DA model (see figure 4.8g) in overall accurate prediction of biogas.

Table 4.4 Performance of the developed ensemble models

Model	ACC	TPR class			TNR class		
		Low	Medium	High	Low	Medium	High
Proposed	**91**	**90**	**88**	**93**	**91**	**92**	**89**
Benchmark stacked models							
Low based	68	**85**	72	53	62	66	**78**
Medium based	60	50	**84**	47	64	48	69
High based	71	50	68	**87**	**78**	72	60
Hard voting	51	30	**76**	43	58	38	56
Soft voting	63	45	56	80	69	66	51
Benchmark bagging models							
RF	75	**83**	72	73	73	**77**	**77**
Subspace NB	74	65	65	**87**	77	78	65
Benchmark boosted models							
XGBoost	77	78	76	77	76	77	**77**
Gentle Boost DA	70	78	48	82	67	**80**	60
RUS Boost GPR	72	55	72	**83**	**78**	72	64

Furthermore, the results of the ROC curve for the best performing ensemble models presented in Table 4.5 showed that the proposed ensemble model outperformed the other benchmark models as it had the highest TPR and the lowest FPR for the three classes of biogas compared to the other benchmark models. The proposed ensemble model demonstrated TPRs above 0.8 for the three classes of biogas compared to the other benchmark models which had lower TPRs. The proposed ensemble model also demonstrated FPRs below 0.15 compared to the other benchmark models which had FPRs from 0.15 and above. The result of the TPRs and FPRs obtained for the proposed ensemble model is another indication confirming the high level of accuracy of the proposed model as it obtained relatively high TPRs (i.e., above 0.8) with a corresponding low FPR (i.e., below 15%).

Generally, the observations made in the results of the performance of the proposed ensemble model obtained in this research study, provide strong evidence of the superior performance of the proposed ensemble model in comparison to the other benchmark models. The superior performance of the proposed ensemble model also confirms the effectiveness of the proposed model in improving the accuracy of biogas predictions from the AD plant. This research finding is in line with achieving the aims and objectives of this research study.

Hence, the developed ensemble model has the potential to improve the performance of AD in the production of biogas. The implication of this indicates that the proposed ensemble model can help to address some of the technical limitations of this research study as it can help to improve the effectiveness of the micro-AD plant in the production of biogas leading to an increase in the production of sustainable biogas. This is a positive approach in contributing towards achieving the full potential of the AD technology which has been revealed in previous research works to have the potential of meeting the world's ever-increasing energy demands through the production of biogas (a renewable alternative energy source) (Abdel daiem & said, 2022; Obaideen *et al.*,2022). Also, the increase in the production of sustainable biogas will have positive impacts on the environment as it plays a distinct role in the ongoing fight against global warming and climate change (Alrowais *et al.*,2023). This is because the production of sustainable biogas replaces dependency on fossil fuels, reduces the energy demand of waste treatment plants, and can yield valuable organic fertilizers useful in improving agricultural yield compared to chemical fertilizers which are still being used in many parts of the world (Pohl *et al.,* 2012)

Table 4.5: Result of the ROC curve of the best performing ensemble models

Model	Parameter	Sample										
		0	1	2	3	4	5	6	7	8	9	10
		High class										
Proposed model	FPR_3	1.00	0.47	0.35	0.18	**0.10**	0.07	0.04	0.03	0.03	0.02	0.01
	TPR_3	1.00	0.99	0.99	0.98	**0.95**	0.92	0.88	0.75	0.63	0.50	0.39
	AOC	0.53	0.12	0.16	0.08	0.03	0.02	0.01	0.00	0.00	0.01	0.00
	G-mean	0.00	0.73	0.80	0.89	**0.93**	0.93	0.92	0.85	0.79	0.70	0.62
Benchmark subspace NB	FPR_3	1.00	0.92	0.89	0.82	0.70	0.57	**0.32**	0.28	0.20	0.11	0.03
	TPR_3	1.00	0.96	0.94	0.93	0.90	0.80	**0.69**	0.62	0.44	0.20	0.08
	AOC	0.08	0.03	0.07	0.11	0.12	0.20	0.03	0.05	0.04	0.02	0.00
	G-mean	0.00	0.28	0.32	0.41	0.52	0.59	**0.68**	0.67	0.59	0.42	0.27
Benchmark RF	FPR_3	1.00	0.91	0.80	0.62	0.53	0.44	**0.21**	0.16	0.11	0.07	0.03
	TPR_3	1.00	0.99	0.96	0.93	0.84	0.78	**0.67**	0.63	0.37	0.21	0.13
	AOC	0.09	0.11	0.17	0.08	0.08	0.18	0.03	0.03	0.01	0.01	0.00
	G-mean	0.00	0.30	0.44	0.59	0.63	0.66	**0.73**	0.73	0.57	0.44	0.36
Benchmark	FPR_3	1.00	0.93	0.77	0.49	**0.38**	0.35	0.29	0.20	0.12	0.04	0.03

XGBoost	TPR_3	1.00	0.98	0.91	0.78	**0.76**	0.68	0.62	0.48	0.37	0.26	0.14
	AOC	0.07	0.16	0.25	0.09	0.02	0.04	0.06	0.04	0.03	0.00	0.00
	G-mean	0.00	0.26	0.46	0.63	**0.69**	0.66	0.66	0.62	0.57	0.50	0.37
						Medium class						
Proposed model	FPR_2	1.00	0.47	0.36	0.27	0.17	0.13	**0.06**	0.05	0.04	0.03	0.01
	TPR_2	1.00	0.96	0.94	0.92	0.89	0.87	**0.81**	0.64	0.49	0.31	0.18
	AOC	0.53	0.10	0.09	0.09	0.03	0.07	0.01	0.00	0.00	0.01	0.00
	G-mean	0.00	0.72	0.78	0.82	0.86	0.87	**0.87**	0.78	0.69	0.55	0.42
Benchmark subspace NB	FPR_2	1.00	0.96	0.79	0.73	0.63	0.55	**0.49**	0.44	0.32	0.24	0.11
	TPR_2	1.00	0.95	0.86	0.83	0.77	0.71	**0.64**	0.55	0.41	0.26	0.17
	AOC	0.04	0.16	0.05	0.08	0.06	0.04	0.03	0.07	0.03	0.03	0.02
	G-mean	0.00	0.19	0.42	0.47	0.53	0.57	**0.57**	0.55	0.53	0.44	0.39
Benchmark RF	FPR_2	1.00	0.93	0.84	0.59	0.50	0.31	**0.26**	0.21	0.11	0.04	0.03
	TPR_2	1.00	0.92	0.84	0.77	0.73	0.66	**0.62**	0.54	0.41	0.29	0.12
	AOC	0.07	0.08	0.21	0.07	0.14	0.03	0.03	0.05	0.03	0.00	0.00
	G-mean	0.00	0.25	0.37	0.56	0.60	0.67	**0.68**	0.65	0.60	0.53	0.34
Benchmark XGBoost	FPR_2	1.00	0.87	0.74	0.65	0.58	**0.52**	0.46	0.41	0.30	0.24	0.14
	TPR_2	1.00	0.93	0.88	0.81	0.75	**0.69**	0.60	0.55	0.43	0.35	0.26
	AOC	0.13	0.12	0.08	0.06	0.05	0.04	0.03	0.06	0.03	0.04	0.04
	G-mean	0.00	0.35	0.48	0.53	0.56	**0.58**	0.57	0.57	0.55	0.52	0.47
						Low class						
Proposed model	FPR_1	1.00	0.47	0.36	0.27	0.17	**0.13**	0.06	0.05	0.04	0.02	0.01
	TPR_1	1.00	0.99	0.97	0.96	0.94	**0.92**	0.83	0.67	0.56	0.33	0.18
	AOC	0.53	0.11	0.09	0.10	0.03	0.07	0.01	0.00	0.01	0.00	0.00
	G-mean	0.00	0.73	0.79	0.84	0.88	**0.89**	0.88	0.80	0.74	0.57	0.42
Benchmark subspace NB	FPR_1	1.00	0.94	0.73	0.45	**0.31**	0.24	0.18	0.17	0.12	0.09	0.02
	TPR_1	1.00	0.99	0.96	0.84	**0.67**	0.58	0.47	0.35	0.31	0.24	0.12
	AOC	0.06	0.21	0.27	0.12	0.05	0.03	0.00	0.02	0.01	0.02	0.00
	G-mean	0.00	0.24	0.51	0.68	**0.68**	0.66	0.62	0.54	0.52	0.47	0.34
Benchmark RF	FPR_1	1.00	0.95	0.75	0.57	0.49	0.42	0.34	**0.26**	0.21	0.14	0.09
	TPR_1	1.00	0.99	0.97	0.96	0.89	0.84	0.74	**0.74**	0.62	0.39	0.23
	AOC	0.05	0.20	0.17	0.08	0.06	0.07	0.06	0.04	0.04	0.02	0.02
	G-mean	0.00	0.22	0.49	0.64	0.67	0.70	0.70	**0.74**	0.70	0.58	0.46
Benchmark XGBoost	FPR_1	1.00	0.93	0.63	0.54	0.48	0.33	0.23	**0.15**	0.07	0.05	0.02
	TPR_1	1.00	0.99	0.95	0.90	0.86	0.74	0.64	**0.54**	0.52	0.30	0.09
	AOC	0.07	0.30	0.09	0.05	0.13	0.07	0.05	0.04	0.01	0.01	0.00
	G-mean	0.00	0.26	0.59	0.64	0.67	0.70	0.70	**0.68**	0.70	0.53	0.30

AOC: Area of curve G-mean: geometric mean FPR_i: False positive rate in i^{th} class TPR_i: True positive rate in i^{th} class

4.11 Outcome of the Uncertainty and Sensitivity Analysis

Figures 4.12(a) and (b) shows the results of the uncertainty and sensitivity analysis for the developed ensemble model for the prediction of biogas. From the results presented in figure 4.12a, it was observed that the models demonstrated a notable ability to be trained using a reduced dataset size of up to 80% of the total available training data, all the while maintaining a remarkable

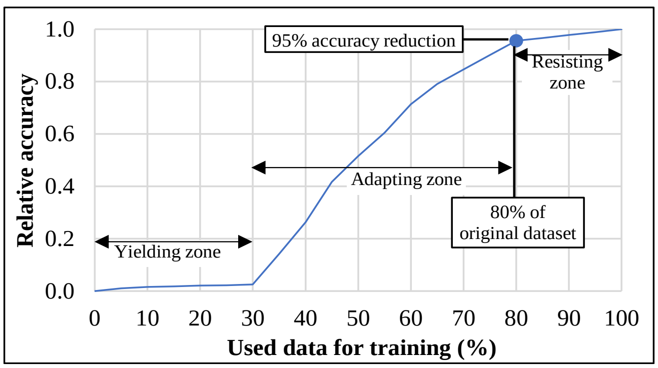
95% accuracy retention. In simpler terms, the models displayed robust performance within this specified range of training dataset size, displaying resilience against the impact of dataset reduction.

However, it is important to note that as the dataset size was further reduced beyond this range, the models exhibited an adaptive behaviour resulting in a gradual nonlinear decline in accuracy. Additionally, as the dataset size decreased to less than 30%, the models encountered significant challenges, which led to a complete failure in performance, with accuracy levels decreasing abruptly to almost 0%. This particular revelation bears substantial significance. Despite the models being developed within a relatively uncomplicated framework utilising input data spanning close to a year, the demonstrated adaptability and efficiency within this context can have far-reaching implications. Such efficiencies have the potential to yield substantial energy cost savings. The effect of this can have positive benefits on the application of AD technology as it will make the AD technology a more economical technology for the management of organic wastes.

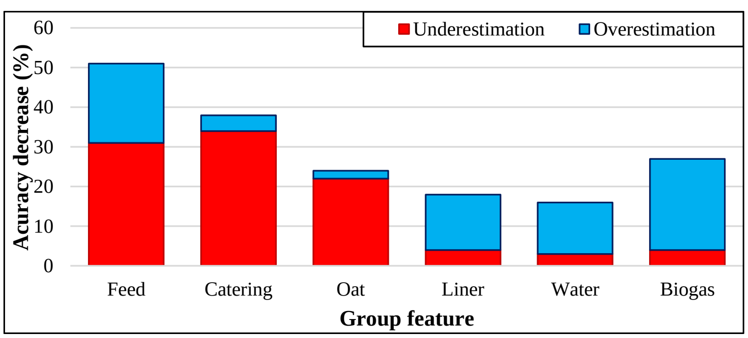
This approach will also make the AD technology more attractive for potential investors/government officials especially in developing countries where the adoption of the AD technology for the management of organic wastes has been discouraged due to high operating and maintenance costs. In addition, it can also mitigate the need for recurrent and time-consuming retraining. This will make it relatively easy to implement especially within the context of broader industrial applications.

The sensitivity analysis focused on the influence of removing individual groups of features as presented in Figure 4.11b. The observed decline in overall accuracy underscores the pivotal role of specific group features, particularly the feeding-related attributes, in shaping the model's performance. Also, the removal of these feed-related features resulted in a substantial 50% drop in accuracy. This contributed significantly to both overestimation and underestimation tendencies. Interestingly, the nature of impact varies across different group features. Specifically, the removal of catering and oat-related group features primarily led to an increase in underestimation, while the attributes related to biogas, water, and liner exerted a more pronounced effect on overestimation tendencies.

This outcome underscores the critical importance of the composition of input materials, with each material potentially exerting a distinct influence on the model's predictive performance. Moreover, insights derived from previous biogas production provide valuable clues to the model regarding the residual potential for the release of biogas. This, in turn, has the potential to mitigate overestimation in future predictions. Remarkably, the incorporation of this group of features as input has the capacity to alleviate a significant portion of overestimation instances, effectively addressing approximately 20% of such cases.



(a)



(b)

Figure 4.12 (a) uncertainty analysis on the dataset (b) sensitivity analysis

4.12 Performance of the Applied Optimisation Method

The SFLA optimisation method was used to specify both the optimal weekly condition and best input pattern for obtaining maximum volume of biogas from the micro-AD plant. The results obtained are presented in Figure 4.13a. This pattern undergoes rigorous testing with previously unobserved data spanning over a period of 76 days. A comparative analysis was then conducted against the actual operational performance which recorded the most productive phase of biogas generation during this period.

Figure 4.13b clearly demonstrates the efficacy of the proposed pattern in increasing the number of days with high biogas yield. According to the data, the implementation of the proposed pattern led to a notable 78% increase in the duration of time during which substantial biogas production was achieved. This significant improvement indicates the effectiveness of the proposed approach in the optimisation of biogas generated from the AD plant. From figure 4.13a, it was observed that high portions of feed feature were fed into the digester only on the 4th and 7th days. The catering feature was fed on the 3rd, 4th, 6th, and 7th days at high and low portions, respectively. Low portions of the oat feature were fed on the 3rd and 6th days. Water was added to the digester on the 3rd, 4th, 6th, and 7th days at low amounts to obtain maximum volume of biogas. Intriguingly, the absence of the liner input in the optimal condition implies it has no significant impact on enhancing biogas generation. This finding underscores that the inclusion of the liner feature did not significantly contribute to the overall biogas yield, thus making its omission from the optimal setup a justifiable decision. This justifiable decision is based on the analysis of the operator which revealed that an insignificant portion of liner was added to the digester. In the case of the feed feature, high-class variables were strategically incorporated into the digester for a mere two days within the week. This observation reveals a potential strategy for the optimisation of biogas generation, indicating that rather than the continual input of materials, a more effective approach could involve extending the intervals between the addition of the feed feature by 2 or 3 days, followed by a substantial surge in the system's load. This noticeable difference becomes apparent in Figure 4.13c, where a clear departure from the usual practice of frequent waste feeding into the pre-digester (black dots) is vividly reduced by implementation of the proposed strategy (red dots). Results highlight a substantial and statistically significant reduction of 71% in the amount of time spent on operational activities, during which the mechanised pumping mechanism facilitates the controlled transfer of materials into the digester. This reduction carries important implications such as energy conservation, as well as a notable decrease in the demands for careful monitoring and extensive maintenance efforts. In addition, the manner in which catering, and oat materials are suggested highlights contrasting patterns. The model proposes an initial infusion of a substantial quantity of catering materials into the pre-digester, followed by a subsequent day with a low catering input. Conversely, a light input of oat material on one day is suggested. Furthermore, the recommended approach for adding water demonstrates a distinct trend. It suggests that the addition of low amounts of water should be prioritised on days when waste materials are input (as evident on days 3, 4, 6 and 7 in Figure4.13a). Upon comparing the input pattern with the real case observations presented in Figure A4 of the Appendix, it becomes evident that the total operation days for each input increased by approximately 25%. However, when considering the overall picture, as presented in Figure 4.13d, a 30% decrease in the total number of pre-feeding days were observed. This implies that despite the individual increase in the number of input materials being added, the strategy of compacting them on specific days contributed to a reduction in the overall operation days and associated costs. The results of the applied optimisation method (SFLA) presented in the figure 4.13 for obtaining maximum biogas generation indicates that the generation of maximum biogas volume from the micro-AD plant can be achieved with minimum labour and minimum energy cost. This is important as it helps to reduce the operational cost of the micro-AD plant while obtaining maximum volume of biogas.

The effect of this can help to address the high operational costs which is one of the economic challenges associated with AD plants.

4.13 Practical Implementation Challenges of the Proposed AI-Based Solutions for New and Existing AD systems.

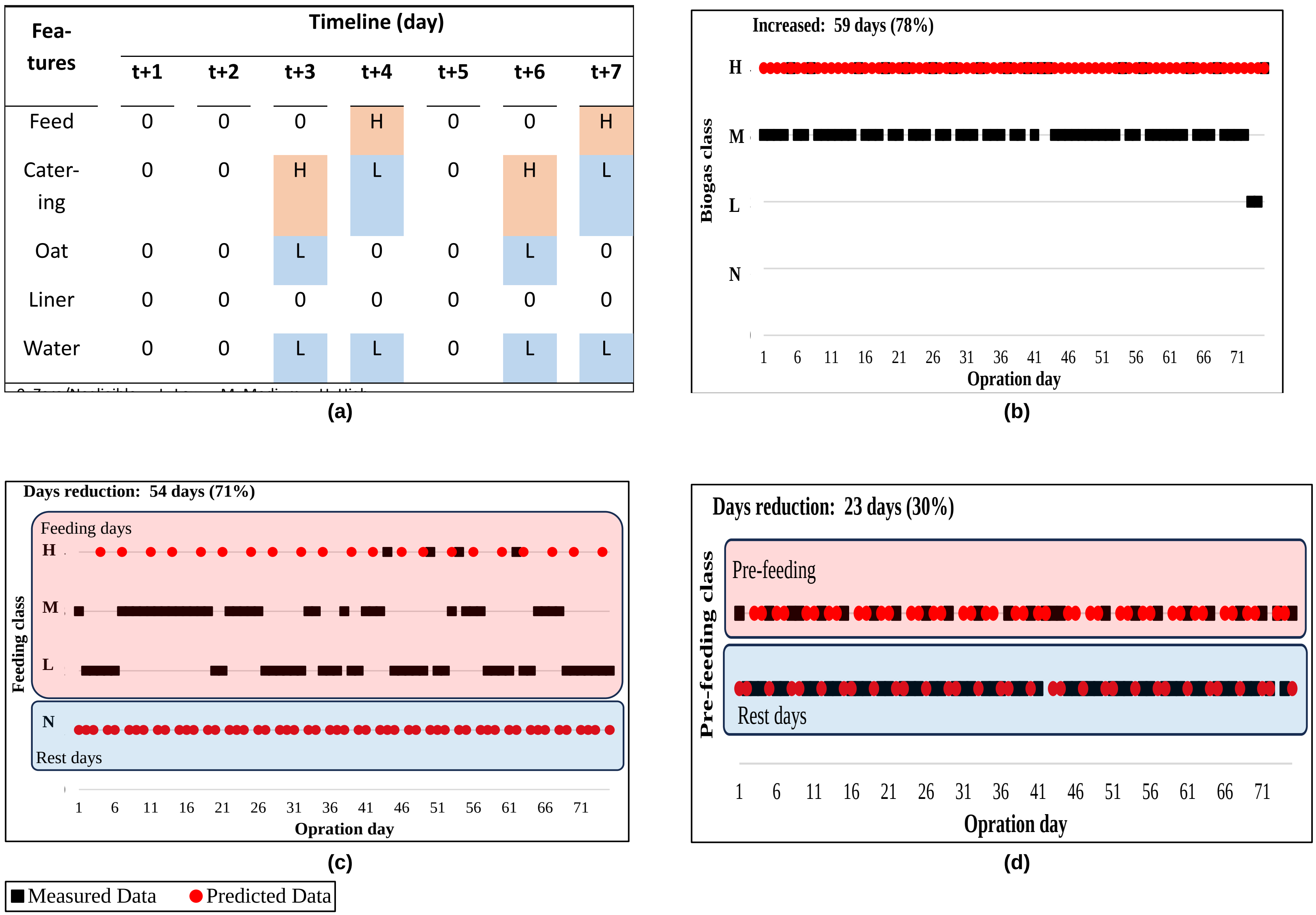
The development of both RNN-SFLA model and time-series ensemble model using historic data obtained from a micro-AD plant provided vital information to the AD operator on how the performance of the micro-AD plant can be improved to generate maximum biogas volume which is highly vital especially for meeting the energy demands of the people. The information provided from these developed models also gave an insight on the optimal operation patterns/strategy which can yield maximum volume of biogas based on the different feedstock used with minimal amount of energy and labour. The information provided based on the results obtained demonstrated to have the potential to address the high cost of operations of the plant. This is a major challenge mostly associated with the application of AD for treating organic wastes. However, despite the potentials of these developed models, some challenges associated with the practical implementation of the developed models in other AD plants were observed.

Firstly, the exclusion of operational parameters like temperature and pH amongst others in the development of the RNN model for biogas prediction which gave rise to signs of instability indicated by the sudden rise in drop in the predicted biogas compared to the measured biogas observed in the biogas predicted from the micro-AD plant along days 4-10, 16-19, 22-25 and 40-43 as presented in figure 4.4, has a significant challenge in the application of the RNN model in other AD systems. The challenge associated with the RNN model application lies in its inability to ensure complete stability when applied in other AD plants to predict biogas. This is due to the exclusion of the temperature parameter which plays a vital role in ensuring the stability of the AD plant for biogas production.

Secondly, the optimal operation strategy/pattern obtained for both the RNN model, and time-series ensemble model was based on the feedstock used for the micro-AD plant. Though it provided useful information to the operator of other AD plants however, a crucial challenge lies in its inability to be applied in other AD plants which have completely different composition of feedstock.

Thirdly, the sensitivity analysis carried out for the developed models gave useful information to the operator of the AD plant on the impact of each input variable/parameter on the generation of biogas. The information provided based on this analysis also has its limitations as the sensitivity analysis carried out by the developed models were based on specific feedstocks which might not be applicable in other AD systems for generating biogas.

Though these outlined challenges have the tendency to limit the practical application/implementation in other AD plants, the proposed AI-based models can be applied in AD plants which have similar qualities with the AD plant used for achieving the purpose of this research study.



Fea-tures	t+1	t+2	t+3	t+4	t+5	t+6	t+7
	Timeline (day)						
Feed	0	0	0	H	0	0	H
Cater-ing	0	0	H	L	0	H	L
Oat	0	0	L	0	0	L	0
Liner	0	0	0	0	0	0	0
Water	0	0	L	L	0	L	L

Figure 4.13. Comparison between optimal weekly operation pattern and testing event (76 days): (a) suggested optimum condition for the operation of the micro-AD plant for maximum biogas generation, (b) yielded biogas, (c) feeding to digestor, and (d) pre-feeding days

5 Chapter 5. Conclusions

5.1 Summary

This research study first developed a three-step AI-based framework for the estimation of biogas generation from a micro-AD plant. Sequel to the development of the AI-based framework, an ensemble-based framework that offers a suggested real-time weekly operation pattern for improving biogas generation from the micro-AD plant was developed. The first step of this research study entailed the collection of raw data from the micro-AD plant, selection of the relevant parameters from the raw data collected, infilling the missing data observed in the data collected for developing data-driven models using different data mining techniques and the development of the RNN/NARX model. The developed RNN model was trained to predict biogas generated from a micro-AD plant. Following this, the RNN model was tuned using the SFLA optimisation model to specify the optimal variables and obtain the optimal biogas generation from the micro-AD plant. The SFLA optimisation model also showed the optimal weekly pattern for the feeding distribution pattern for the generation of maximum volume of biogas from the AD plant. Sensitivity analysis was conducted on the developed RNN model to determine the impact of each input variable on the volume of biogas generated. Uncertainty analysis was also carried out to determine the correlation between the dataset and the developed model. The second step of this research study analysed the concept behind the extraction and selection of features from the collected data, the development of different WLDM models and the development of the ensemble model from the combination of the different WLDM models. It also emphasized on the use of applied data cubes and data warehouse which assisted in constructing the real-time platform and the KPI performance of WLDMs for developing ensemble models. In addition, the effectiveness of the developed ensemble model in improving the prediction accuracy of biogas was determined through comparisons with different benchmark models developed alongside with the developed ensemble model.

Furthermore, the developed ensemble model was optimised using the SFLA optimisation model to determine the optimal weekly operation for obtaining maximum biogas volume. Similar to the RNN model, further analyses were also carried out on the developed ensemble model. These analyses include sensitivity analysis and uncertainty analysis. The sensitivity analysis was carried out to show the impact of each input feature on the biogas generated from the AD plant. The uncertainty analysis on other hand, was conducted to show the variations in the relative accuracy with the corresponding reduction in the dataset. The development of both model frameworks shows the advances made towards further improving the effectiveness of anaerobic digestion operations for generating maximum volume of biogas. It also shows the efforts made towards tackling the issues associated with the management of organic wastes thereby addressing one of the key environmental challenges being faced in the world today.

5.2 Key findings

Based on the application of the proposed model frameworks for achieving the aims and objectives of this research study, different key findings were observed which provided answers to the research questions raised in the introduction. These key findings are as follows,

- The Kriging technique demonstrated to be the most effective technique for infilling the missing data in the collected dataset compared to the popularly applied conventional techniques, particularly KNN, SVM, LR (linear regression), FFNN and linear interpolation. This indicates that the application of the Kriging technique for infilling missing data has the potential to address the major challenge being faced by many real industrial practices suffering from the lack of proper databases and some inputs such as biogas generation which are not being measured daily as required due to unavoidable constraints. It also indicates that the Kriging technique can contribute to a higher level of model prediction accuracy compared to other conventional infilling techniques applied in previous research studies. The Kriging technique can also play a vital role in improving the optimisation ability of the model with the aim of improving the overall performance of the AD plant. The performance of the kriging technique used to infill the missing data for the development of both RNN and the ensemble model contributed towards achieving objectives 1, 2 3 and 4. It was also a preliminary approach towards addressing research questions 1, 2, 3 and 4 respectively.

- The recommended model tunning method for including data of influential days, i.e., using cross-correlation analysis for finding the lag time could either prevent underestimation by ignoring some influential days or overestimation by including ineffective days. The results obtained show that yielded biogas is highly sensitive to the three waste compositions (i.e., catering, oats, and liner) added to the pre digester as it yielded biogas after one day, whereas it changed gradually to three days for feed and five days for added water. In addition, the model tuning method demonstrated to have the potential to improve the overall efficiency of the AD plant by speeding up the modelling process while improving the accuracy of the model. The ability of the model tuning method to improve the overall efficiency of the AD plant showed a significant attempt made in achieving one of the aims of this research study.
- The developed RNN model had a low average RMSE of 0.39 and a relative value of 2% for 310 datasets. This shows how effective the developed RNN model is in the accurate prediction of biogas even with smaller datasets. The ability of the RNN model to effectively predict biogas from the AD plant was in line with achieving objective 1. However, the developed model also has an NNSE of 0.84 which is acceptable though low compared to other studies. Hence, it requires further investigation.
- The generation of biogas from the micro-AD plant was observed to be strongly influenced by the oats and catering composition compared to the feed, liner composition and water

added to the digester. This finding also confirms the accuracy of the developed RNN model as the analysis of different feedstock carried out by the operator of the plant revealed that the oats and catering composition had a higher influence on the volume of biogas generated compared to the other feedstock. Also, the addition of liner to the micro-AD system can significantly reduce the volume of biogas produced. Furthermore, changing the calendar pattern of adding water, catering, and oat had no significant improvement in the volume of biogas generated. The result performance of the RNN model in the prediction of biogas shows an attempt made to achieve the 1st objective as the effectiveness of the developed RNN model was ascertained using the sensitivity analysis which was compared with the operator's analysis. It also tends to answer the 1st research question as the performance of the RNN model provided information which can be useful to the operator of similar AD systems as regards to its sensitivity to the different input variables in generating biogas.

- The developed optimal operation pattern could result in 15% and 43% increase for a 7-day and 47-day period, respectively in the biogas generation from the micro- AD plant compared to business as usual. The result of the developed optimal operation pattern provided an insight on the potentiality of the optimisation method used in this research study which can be applied in other AD systems having similar feed composition variable with the aim of improving the performance of the AD system in the generation of biogas. These results provided answers to the 2nd research question. The results obtained demonstrated efforts made towards achieving the 2nd objective of this research study which implied that the developed AI based model (RNN-SFLA) has the potential to improve the performance of the micro-AD plant in generating maximum volume of biogas.
- The performance of the developed ensemble model demonstrated to have the highest level of accuracy in giving correct estimations compared to other benchmark models. It demonstrated to have an accuracy of 91% while other benchmark models demonstrated

accuracies observed to be within the range of 50%- 80%. The developed ensemble model also had the least degree of underestimations and over estimations compared to the other developed models. The performance of the developed ensemble model indicates that the ensemble model has the potential to improve the accuracy of biogas prediction thereby serving as a useful tool/information to the operator of the AD system especially on the possible volume of biogas expected from future AD operations. This tends to answer the 3rd research question as it demonstrated its significance in providing vital information useful to the AD operator especially in terms of amount of energy generation and planning. The various comparisons between the developed ensemble model and other benchmark models were targeted at achieving the 3rd objective of this study.

- Both PLS and sequential sensitivity analysis carried out reveal a high sensitivity to the feed feature compared to the other input features.
- . The optimised weekly AD operation demonstrated promising results which was targeted towards achieving the 4th objective and providing answers to the 4th research question. It had a substantial 78% increase in the number of days achieving high volume of biogas generation, accompanied by a 71% reduction in total required feeding days and a 30% reduction in pre-feeding days. These results have a positive economic impact on the overall operation of the AD plant as an increase in the volume of biogas obtained from the AD plant with a significant reduction in the required feeding and pre-feeding days will lead to a corresponding reduction in O & M costs.
- The optimal condition for the weekly operation of AD showed that the liner feature had no influence in obtaining maximum volume of biogas from the micro-AD plant as there was no input from it in the optimization of biogas. Also, the optimal condition revealed that the feed feature was the only feature which was required to be added to the digester only twice a week (i.e., on the 4th and the 7th day) to obtain maximum volume of biogas. This implies that the developed ensemble model can help to reduce operating and

maintenance (O & M) costs of this micro-AD plant while generating maximum volume of biogas. The ability of the developed ensemble model to reduce the O & M cost of AD makes it a more economical technology thereby addressing the 4th research question.

The optimal condition also revealed that no input variable is required to be added to the digester on the 1st, 2^{nd,} and 5th days respectively for maximum volume of biogas to be obtained. In addition, high portion of catering and low portion of oat together with a low amount of water are required on the 3rd day. On the 4th day of the operation, low amounts of water, a low portion of oat, a low portion of catering and a high portion of feed are required. On the 6th day of the operation, a high portion of catering, a low portion of oats and a low amount of water are required to be added to the digester to obtain maximum volume of biogas. Lastly, a low portion of catering composition and a low amount of water are required to be added to the digester in addition to a high portion of feed for obtaining maximum volume of biogas. The optimal feeding pattern implies that the application of the ensemble model can help to overcome some of the economic constraints of the micro-AD plant as the optimal feeding pattern will help to conserve both the energy required for AD operations and the cost of labour while generating maximum volume of biogas. This also provided answers to the 4th research question as well as the 4th research objective.

5.3 Key Contributions to Knowledge and Relevance to the Discipline

The development of both the RNN model and the time-series ensemble model for the real-time operation of a micro- AD plant is a significant advancement in enhancing the predictive capabilities of AI based models for improving the performance of AD operations in generating maximum volume of biogas. This significant advancement is due to the increasing demand to expand the application of AI models in AD systems for classification and regression purposes especially as an initial step before the use of advanced models. Based on this, the application of RNN and the time series ensemble model contributed to knowledge in different ways.

First, the development of the RNN-SFLA model for predicting biogas generation from a micro-AD plant and optimising the generated biogas provided an insight on the effect of different waste compositions and water added to the digester on maximum biogas generation. It also explored the feasibility of RNN model application within the industrial context, taking into consideration earlier timesteps. This is particularly important as previous research works have either used simple ML or ANN models on a laboratory scale which limited its widespread deployment.

Secondly, the development of a time series ensemble model offered a more straightforward approach which can enable AD operators to easily understand and interpret the input and output variable (biogas). It also enabled them to interpret the different classes of biogas. The simplicity of its operation made it more accessible and user-friendly, especially for operators who may not have extensive technical expertise in advanced modelling techniques.

Through the application of these developed models in a micro-AD plant, it can help to simplify the decision-making process as dealing with numbers and volumes in a practical setting can be challenging and cumbersome. They provide clear indications of the operational state of AD system as well as the class to which each input or output variable belong. Moreover, these frameworks introduce user-friendly weekly operation patterns which enable easy implementation by operators which can be applicable in other worldwide micro- AD projects. These weekly operation patterns can also serve as a guide to operators of non-micro-AD projects on how best to efficiently utilise feedstock in obtaining maximum biogas volume.

The key contributions of this research to knowledge hold significant relevance not only to the field of AI but also in the field of organic waste management using AD technology. This is because through the application of these models in AD, some of the technical and economic constraints associated with the AD plant can be addressed. In addition, these models can be applied in similar AD projects to address similar constraints. This will help to further promote the application of AD as an organic waste management technique, especially in many developing countries around the world where the application of AD is yet to be implemented mainly due to some of the technical and infrastructure constraints associated with AD systems. For instance, the optimal weekly pattern helps to minimise the need for enormous amounts of water required for wet AD operations which is a major challenge in developing countries due to shortage in water supply. The optimal weekly operation pattern also helps to reduce the operational cost of labor required for the operation of AD. This is very vital as it is one of the economic constraints currently hindering the implementation of AD in many developing countries. This approach will assist in contributing towards the implementation of circular economy. Furthermore, promoting the application of AD can help to promote net zero emissions into the atmosphere thereby contributing towards climate change mitigation alongside promoting energy security through biogas generation.

5.4 Recommendations for Future Work

The developed RNN and ensemble-based models demonstrated to be effective in predicting and optimizing biogas from the micro-AD plant. This conforms with the aims and objectives of this research study thereby contributing to a gap in knowledge. Both the developed RNN and ensemble models have also shown to have promising potentials in improving the effectiveness of AD in producing biogas. This will not only enable the AD technology to maximise its full potential in meeting up with the world's ever increasing energy demands but it will also assist in contributing both directly and indirectly towards the achievement of the various SDGs such as SDG 1(No poverty), SDG 2 (Zero Hunger), SDG 3 (Good Health and Wellbeing), SDG 4 (Quality Education), SDG 5 (Gender Inequality), SDG 6 (Clean Sanitation and Water), SDG 7 (Clean and Renewable Energy), SDG 8 (Decent Work and Economic Growth), SDG 9(Build Resilient Infrastructure, Promote Sustainable Industrialisation and Foster Innovation), SDG, 11 (Sustainable Cities and Communities), SDG 12 (Sustainable Waste Management), SDG 13 (Climate Action), SDG 14 (Life below Water), SDG 15 (Life on Land) and SDG 16 (Promote peaceful and inclusive societies) through the increased application of AD in many parts of the world where the AD technology is yet to be applied. However, there are certain limitations observed in the course of this research study which need to be acknowledged, as a means of pointing towards avenues for further research and development (R & D).

- For the optimal input pattern, it is recommended that the micro-AD plant is fed in one day and allowed to rest for three days in comparison with gradual feeding to obtain the maximum volume of biogas generation.
- Further research studies are required to be conducted to improve the ability of the developed RNN model in tracking sudden drops or rising biogas, especially because many obstacles in operation may happen in practice which might affect the overall output. Also, the findings obtained in this research study require to be further assessed and verified in other AD plants with longer analysis periods to show the efficacy of the developed RNN/NARX model.
- Further analysis and data modelling are required to address other technical challenges associated with AD technology such as long retention time.
- A series of pre- processing steps such as utilising data mining techniques and supporting the capabilities of remote sensing should be taken into consideration to overcome the limitation of having access to comprehensive big data and operational databases for time spans shorter than a day and datasets spanning over a year.
- The proposed ensemble model and the distinct weekly pattern should be subjected to longer-term analysis and testing across different periods and within comparable AD projects. This will provide a meticulous understanding of the effectiveness of the proposed ensemble model and potential scalability in the prediction and optimisation of biogas.
- An intriguing avenue for exploration involves the integration of the real-time operational pattern with risk scenarios. This avenue could include scenarios such as shifts in the

composition of waste, or errors made by operators while adding input materials which have led to uncertainties in the dataset. The introduced pattern and optimisation framework have the potential to dynamically adapt the weekly pattern to tackle these operational challenges. This approach proffers suggestions for a wider application within the realm of digital visualization projects.

- It is recommended that the impact of both the developed RNN and time-series ensemble model on the composition of biogas should be investigated in future research studies. This is to further ascertain their feasibility in improving the composition of biogas generated from AD. Also, investigations on the limitations of heavy metals mobility on the overall digestate quality should are also recommended for future research works as this research centred mainly on improving the volume of biogas produced from AD.

6 References

Abdel daiem, M., Hatata, A., Said.N., Modelling, and optimization of semi-continuous anaerobic co-digestion of activated sludge and wheat straw using Nonlinear Autoregressive Exogenous neural network and seagull algorithm. *Energy*, 241 (2022), Article 122939

Abdel daiem, M.M., N. Said. Energetic, economic, and environmental perspectives of power generation from residual biomass in Saudi Arabia. Alex. Eng. J., 61 (2022), pp. 3351-3364

Abdul Aziz NIH, Hanafiah MM, Mohamed Ali MY. Sustainable biogas production from agrowaste and effluents-A promising step for small-scale industry income. *Renewable Energy* 2019; 132:363–9

Adnan AI, Ong MY, Nomanbhay S, Chew KW, Show PL. Technologies for biogas upgrading to biomethane: A review. *Bioengineering* 2019; 6:92

Africa Biogas Partnership Programme, 2017. http://africabiogas.org/

Ahmad M, Hengyi H, Rahman ZU, Khan ZU, Khan S, Khan Z. Carbon emissions, energy use, gross domestic product, and total population in China. Ekonomia i Srodowisko. Nr; 2018. p. 32–44.

Akhiar, A., Zamri, M. F. M. A., Torrijos, M., Battimelli, A., Roslan, E., Marzuki, M. H. M., & Carrere, H. (2020, April). Anaerobic digestion industries progress throughout the world. In *IOP Conference Series: Earth and Environmental Science* (Vol. 476, No. 1, p. 012074). IOP Publishing.

Alrawashdeh, K.A.B., Al-Samrraie, L., Al Issa, H.A., Qasem, I., Hussien, A., Al-Zboon, K., 2022. Prediction and optimization of biogas production from OMW digestion using fenton pre-treatment process with particle swarm optimization. *International Journal Des., Nature Ecodynamics*. 17 (2), 157–168

Alejo, L., Atkinson, J., Guzmán-Fierro, V., Roeckel. M., Effluent composition prediction of a two-stage anaerobic digestion process: machine learning and stoichiometry techniques. *Environmental Science Pollution Research*., 25 (21) (2018), pp. 21149-21163

Al, R., Behera, C.R., Zubov, A., Gernaey, K.V., Sin, G., 2019. Meta-modelling based. efficient global sensitivity analysis for wastewater treatment plants–An application to the BSM2 model. *Computing. Chem. Eng*. 127, 233–246.

Almomani, F., 2020. Prediction of biogas production from chemically treated co-digested agricultural waste using artificial neural network. *Fuel* 280, 118573

Alrowais.R., Said Noha, Al-Otaibi.A., Hatata. Y.A., Essa. A.M., Abdel daiem. M.M (2023). Comparing the effect of mesophilic and thermophilic anaerobic co-digestion for sustainable biogas production: An experimental and recurrent neural network model study. *Journal of Cleaner Production* 392 (2023), 136248

Alzubi J., Nayyar A., Kumar A., (2018). Machine Learning from Theory to Algorithms: An Overview. IOP Conf. Series: *Journal of Physics: Conference. Series* 1142 (2018) 012012, doi:10.1088/1742-6596/1142/1/0120.

Antwi, P., Li, J., Boadi, P.O., Meng, J., Shi, E., Deng, K., Bondinuba, F.K., 2017. Estimation of biogas and methane yields in an UASB treating potato starch processing wastewater with backpropagation artificial neural network. *Bioresource. Technology*. 228, 106–115.

Ardolino, F., Cardamone, G.F., Parrillo, F., Arena, U., 2021. Biogas-to-biomethane upgrading: A comparative review and assessment in a life cycle perspective. *Renewable. Sustainable. Energy*. Rev. 139, 11058

Arthurson V. Closing the global energy and nutrient cycles through application of biogas residue to agricultural land-potential benefits and drawbacks. *Energies* 2009; 2:226–42.

Arif Sania, Liaquadt Rabia, Manal Adil. (2018). Applications of materials as additives in anaerobic digestion technology. Renewable Sustainable Energy Review 97, 354-366**.**

Arun, C., Sivashanmugam. P. (2017). Study on optimization of process parameters for enhancing the multi-hydrolytic enzyme activity in garbage enzyme produced from preconsumer organic waste. *Bioresource Technology* 226, pp. 200-210

Asgari, S., Gupta, R., Puri, I.K., Zheng, R., 2021a. A data-driven approach to simultaneous fault detection and diagnosis in data centers. *Appl. Soft Comput*. 110, 107638

Asadi, M., McPhedran, K., 2021. Biogas maximization using data-driven modelling with uncertainty analysis and genetic algorithm for municipal wastewater anaerobic digestion. *Journal of Environmental Management.* 293, 112875.

Asgari, S., Moazamigoodarzi, H., Tsai, P.J., Pal, S., Zheng, R., Badawy, G., Puri, I.K., 2021c. Hybrid surrogate model for online temperature and pressure predictions in data centres. *Future. Generation. Computing. System*. 114, 531–547

Ajayi-Banji, A., Rahman, S., 2022. A review of process parameters influences in solid-state anaerobic digestion: Focus on performance stability thresholds*. Renewable. Sustainable. Energy Rev*. 167, 112756.

Auer, A., Vande Burgt, N.H., Abram, F., Barry, G., Fenton, O., Markey, B.K., Nolan, S., Richards, K.G., Bolton, D.J., de Waal, T., Gordon, S.V., O'Flaherty, V., Whyte, P., & Zintl, A. (2017). Agricultural anaerobic digestion power plants in Ireland and Germany: policy and practice. *Journal of the science of food and agriculture, 97 3*, 719-723.

Austin, G., Morris, G. (2012). Biogas Production in Africa. In: Janssen, R., Rutz, D. (eds) *Bioenergy for Sustainable Development in Africa*. Springer, Dordrecht.

Awhangbo, L., Bendoula, R., Roger, J. M., & Béline, F. (2020). Multi-block SO-PLS approach based on infrared spectroscopy for anaerobic digestion process monitoring. *Chemometrics and Intelligent Laboratory Systems*, *196*, 103905.

Azeem K., Muhammad A., Muzammil A., Tariq M., Lorna D. The anaerobic digestion of solid organic waste: A Review. *Journal of Waste Management* 31 (2011), 1737-1744

Baek, G., Lee, C., & Yoon, J. (2023). Machine learning approach for predicting anaerobic digestion performance and stability in direct interspecies electron transfer-stimulated environments. *Biochemical Engineering Journal*, 193, 108840.

Barahmand, Z., & Samarakoon, G. (2022). Sensitivity Analysis and Anaerobic Digestion Modeling: A Scoping Review. *Fermentation*, *8*(11), 624.

Barik, D., Murugan, S., An Artificial Neural Network and Genetic Algorithm Optimized Model for Biogas Production from Co-digestion of Seed Cake of Karanja and Cattle Dung

Beltramo, T., Klocke, M., Hitzmann, B., 2019. Prediction of the biogas production using GA and ACO input features selection method for ANN model. *Inf. Process. Agric*. 6, 349–35

Boe, K., Batstone, D. J., Steyer, J. P., & Angelidaki, I. (2010). State indicators for monitoring the anaerobic digestion process. *Water research*, *44*(20), 5973-5980.

Bong, C.P.C., Ho, W.S., Hashim, H., Lim, J.S., Ho, C.S., Peng Tan, W.S., Lee. C.T., (2017). Review on the renewable energy and solid waste management policies towards biogas development in Malaysia. *Renewable Sustainable Energy* Review 70, 988-998.

Boulamanti, A.K., Donida Maglio, S., Giuntoli, J., Agostini, A., 2013. Influence of different practices on biogas sustainability. *Biomass Bioenergy* 53, 149–161. https:// doi.org/10.1016/j.biombioe.2013.02.020

Boulesteix. A. & Strimmer. K. 2007. Partial Least Squares: A Versatile Tool for the Analysis of High-Dimensional Genomic Data. February 2007. Briefings in Bioinformatics 8(1):32-44.

Bouallagui, H., Rachdi, B., Gannoun, H., Hamdi, M., 2009b. Mesophilic and thermophilic anaerobic co-digestion of abattoir wastewater and fruit and vegetable waste in anaerobic sequencing batch reactors. *Biodegradation* 20, 401–409

Braganca I, Sanchez-Soberon F, Pantuzza GF, Alves A, Ratola N. Impurities in biogas: analytical strategies, occurrence, effects, and removal technologies. Biomass Bioenergy 2020; 143:105878

Brambilla, M., Araldi, F., Marchesi, M., Bertazzoni, B., Zagni, M., Navarotto. P., Monitoring of the startup phase of one continuous anaerobic digester at pilot scale level. *Biomass and Bioenergy*, Volume 36, 2012, Pages 439-446, ISSN 0961-9534,

Castillo, E.F.M., Cristancho, D.E., Arellano, V.A., 2006. Study of the operational conditions for anaerobic digestion of urban solid wastes. *Waste Management*. 26, 546–556

Chae, K.J., Jang, A., Yim, S.K., Kim, I.S., 2008. The effects of digestion temperature and temperature shock on the biogas yields from the mesophilic anaerobic digestion of swine manure. *Bioresource. Technology*. 99, 1–6

Charles, W., Walker, L., Cord-Ruwisch, R., 2009. Effect of pre-aeration and inoculum on the start-up of batch thermophilic anaerobic digestion of municipal solid waste. *Bioresource. Technology*. 100, 2329–2335.

Chauhan, A. (2021). Ensemble methods-bagging, boosting, and stacking.

Cheela V., Ranjan.V., Goel.S., John D.B., (2021). Pathways to sustainable waste management in Indian Smart Cities. *Journal of Urban Management*., 10 (4) (2021), pp. 419-429

Cheon, A. J., Sung, H. Jun, H. Jang, M. Kim, J. Park. Application of Various Machine Learning Models for Process Stability of Bio-Electrochemical Anaerobic Digestion *Processes*, 10 (1) (2022), p. 158

Chicco, D., Tötsch, N., Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *Biodata Mining*, 14(13), pp. 1-22

Chiu, M.C., Wen, C.Y., Hsu, H.W., Wang, W.C., 2022. Key wastes selection and prediction improvement for biogas production through hybrid machine learning methods. *Sustainable Energy Technology Assessments* 52, 102223

Choi, S., Kim, S.I., Yulisa, A., Aghasa, A., Hwang. S., Proactive Prediction of Total Volatile Fatty Acids Concentration in Multiple Full-Scale Food Waste Anaerobic Digestion Systems Using Substrate Characteristics with Machine Learning and Feature Analysis *Waste Biomass Valorization*. (2022), pp. 1-16

Chozhavendhan, S., Karthigadevi G., Bharathiraja B., Kumar R., Abo, L., Prabhu, S., Balachandar, R., Jayakumar, M., Current and prognostic overview on the strategic exploitation of anaerobic digestion and digestate: A review *Environ. Res*., 216 (2) (2023), Article 114526

Cinar, S.O., Cinar.S., Kutcha.K. Machine Learning Algorithms for Temperature Management in the Anaerobic Digestion Process. *Fermentation*, 8 (2) (2022), p. 65

Cruz I., W. Chuenchart, F. Long, K. Surendra, L. Andrade, M. Bilal, H. Liu, R. Figueiredo, S. Khanal, L. Ferreira. Application of machine learning in anaerobic digestion: Perspectives and challenges. *Bioresources. Technology*., 345 (2022), Article 126433

Dahunsi, S.O., Oranusi, S., Efeovbokhan, V.E., 2017. Cleaner energy for cleaner production: modeling and optimization of biogas generation from Carica papayas (Pawpaw) fruit peels. *Journal of Cleaner Production*. 156, 19–29.

De Clercq, D., Jalota, D., Shang, R., Ni, K., Zhang, Z., Khan, A., ... & Yuan, K. (2019). Machine learning powered software for accurate prediction of biogas production: A case study on industrial-scale Chinese production data. *Journal of cleaner production*, *218*, 390-399.

De Clercq, D., Wen, Z., Fei, F., Caicedo, L., Yuan, K., Shang.R., Interpretable machine learning for predicting biomethane production in industrial-scale anaerobic co-digestion.
Science of Total Environment., 712 (2020)

Dela-Rubia, M.A., Perez, M., Romero, L.I., Sales, D., 2002. Anaerobic mesophilic and thermophilic municipal sludge digestion. *Chem. Biochemistry. Eng. Qual*. 16, 119– 124.

Deublein D, Steinhauser A. Biogas from waste and renewable resources: an introduction, John Wiley & Sons; 2011.

Dhussa, A.K., Sambi, S.S., Kumar, Shashi, Kumar, Sandeep, Kumar, Surendra, 2014. Nonlinear Autoregressive Exogenous modeling of a large anaerobic digester producing biogas from cattle waste. *Bioresource. Technology*. 170, 342–349

Dixon, M., Gallop, J. R., Lambert, S. C., & Healy, J. V. (2007). Experience with data mining for the anaerobic wastewater treatment process. *Environmental Modelling & Software*, *22*(3), 315-322.

Dong, C., & Chen, J. (2019). Optimisation of process parameters for anaerobic fermentation of corn stalk based on least squares support vector machine. *Bioresource technology*, *271*, 174-181.

Dos Santos, L.A., Valenca, R.B., da Silva, L.C.S., de Barros Holanda, S.H., da Silva, A.F.V., Juca, ´ J.F.T., Santos, A.F.M.S., 2020. Methane generation potential through anaerobic digestion of fruit waste. *Journal of Cleaner Production*. 256, 120389.

Edwards, J., Othman, M., Burn, S., 2015. A review of policy drivers and barriers for the use of anaerobic digestion in Europe, the United States and Australia. *Renewable. Sustainable. Ener*gy. Review. 52, 815–828. https://doi.org/10.1016/j.rser.2015.07.112

Eghbali, A.H., Behzadian, K., Hooshyaripor, F., Farmani, R., Duncan. A.P., Improving prediction of dam failure peak outflow using neuroevolutionary combined with K-means clustering. *Journal of Hydrological Engineering*., 22 (6) (2017)

El-Mashad, H.M., Wilko, K.P., Loon, V., Zeeman, G., 2003. A model of solar energy utilisation in the anaerobic digestion of cattle manure. *Biosystem. Engineering*. 84, 231– 238.

Eurostat, European Statistics, 2017. http://ec.europa.eu/eurostat (accessed July 1, 2017)

Eusuff.M., Lansey.K., & Pasha. F. (2006) Shuffled frog-leaping algorithm: a memetic meta-heuristic for discrete optimization, *Engineering Optimization*, 38:2, 129-154, Ewusi, A., Ahenkorah, I., & Aikins, D. (2021). Modelling of total dissolved solids in water supply systems using regression and supervised machine learning approaches. *Applied Water Science*, 11(2), 1-16.

Ezekoye, V. A., Ezekoye, B. A., & Offor, P. O. (2011). Effect of retention time on biogas production from poultry droppings and cassava peels. *Nigerian Journal of Biotechnology*, *22*, 53-59.

Fairfield (2021). Anaerobic Digestion Market Value to Witness 17.7% Expansion Through 2026, Foresees a New Study of Fairfield Market Research. Available at www.fairfieldmarket-research.com. Accessed December 2023.

Fazzo, L., De Santis, M., Beccaloni, E., Scaini, F., Iavarone, I., Comba, P., Airoma, D. (2020). A Geographic Information System-Based Indicator of Waste Risk to Investigate the Health Impact of Landfills and Uncontrolled Dumping Sites. *International Journal of. Environmental Research Public Health 17*(16), 5789.

Feng, S., Zhou, H., Dong.H., (2019) Using deep neural networks with small datasets to predict material defects. *Materials and Design*., 162 (2019), pp. 300-310

Fernandez, J., Perez, M., Romero, L.I., 2008. Effect of substrate concentration on dry mesophilic anaerobic digestion of organic fraction of municipal solid waste (OFMSW). *Bioresource. Technology*. 99, 6075–6080

Fezzani, B., Cheikh, R.B., 2010. Two-phase anaerobic co-digestion of olive mill wastes in semi-continuous digesters at mesophilic temperature. *Bioresource. Technology*. 101, 1628–1634.

Formowitz, B., Fritz, M., 2010. Biogas digestates as organic fertilizer in different crop rotations, 18th European biomass conference exhibition. Lyon, France. https://doi.org/10.5071/18thEUBCE2010-OD7.2.

Gautam, R. Baral, S. and Herat, S. 2009. Biogas as a sustainable energy source in Nepal: Present status and future challenges. *Renewable and Sustainable Energy Reviews* 13: 248-252.

Gharib, A., Davies, E. (2021). A workflow to address pitfalls and challenges in applying machine learning models to hydrology. *Advances in Water Resources*, 152, p.103920

Ghatak, M. Das, Ghatak, A., 2018. Artificial neural network model to predict behavior of biogas production curve from mixed lignocellulosic co-substrates. Fuel 232, 178–189.

Gijzen, H.J. (2002). "Anaerobic digestion for sustainable development: a natural approach". *Water Science and Technology*. 45 (10): 321–328. doi:10.2166/wst.2002.0364. PMID 12188565.

Guo, H., Jeong, K., Lim, J., Jo, J., Kim, Y. M., Park, J. P., ... & Cho, K. H. (2015). Prediction of effluent concentration in a wastewater treatment plant using machine learning models. *Journal of Environmental Sciences*, *32*, 90-101

Guo, H., Wu, S. Biao, Tian, Y. Jie, Zhang, J., Liu, H. Tao, 2021. Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: A review. *Bioresources. Technology*. 319, 124114.

Guo, Y., Tian, X., Fang, G., Xu, Y., 2020. Many-objective optimization with improved shuffled frog leaping algorithm for inter-basin water transfers, *Advances in Water Resources*, Volume 138, 103531, ISSN 0309-1708

Gupta, A., Jain, V., & Singh, A. (2022). Stacking ensemble-based intelligent machine learning model for predicting post-COVID-19 complications. *New Generation Computing*, *40*(4), 987-1007.

Gupta R., Zhang.L., Hou.J., Zhang.Z., Liu H., You, S., Ok. Y.S., Li.W., Review of explainable machine learning for anaerobic digestion. Review. *Bioresource Technology* 369 (2023), 128468.

Guwy, A. J., F. R. Hawkes, S. J. Wilcox, and D. L. Hawkes. 1997. Neural network and on-off control of bicarbonate alkalinity in a fluidized-bed anaerobic digester. *Water Research* 31 (8):2019–25.

Hachicha, W., & Ghorbel, A. (2012). A survey of control-chart pattern-recognition literature (1991–2010) based on a new conceptual classification scheme. *Computers & Industrial Engineering*, *63*(1), 204-222.

Haffiez, N., Chung, T.H., Zakaria, B.S., Shahidi, M., Mezbahuddin S., Maal-Bared, R., Dhar. B.R., Exploration of machine learning algorithms for predicting the changes in abundance of antibiotic resistance genes in anaerobic digestion. *Science of Total Environment*., 156211 (2022)

Hamed, M. M., Khalafallah, M. G., & Hassanien, E. A. (2004). Prediction of wastewater treatment plant performance using artificial neural networks. *Environmental Modelling & Software*, *19*(10), 919-928.

Hansupalak N, Piromkraipak P, Tamthirat P, Manitsorasak A, Sriroth K, Tran T. Biogas reduces the carbon footprint of cassava starch: a comparative assessment with fuel oil. *Journal of Cleaner Production* 2016; 134:539–46.

Herbes, C., Roth, U., Wulf, S., Dahlin, J., 2020. Economic assessment of different biogas digestate processing technologies: A scenario-based analysis. *Journal of Cleaner Production*. 255, 120282.

Herrmann A. Biogas production from maize: current state, challenges, and prospects. 2. Agronomic and environmental aspects. *Bioenergy Resources* 2013; 6:372–87

Holubar, P., L. Zani, M. Hager, W. Froschl, Z. Radak, 2002. Advanced controlling of anaerobic digestion by means of hierarchical neural networks. *Water Resources*., Issue 36, pp. 2582-258.

Holubar, P., L. Zani, M. Hager, W. Fröschl, Z. Radak, and R. Braun. 2003. Start-up and recovery of a biogas-reactor using a hierarchical neural network-based control tool. *Journal of Chemical Technology & Biotechnology* 78 (8):847–54. doi:10.1002/jctb.854

Hoppe, T., Sander. M.P.T., (2014). Agricultural green gas demonstration projects in The Netherlands. A stakeholder analysis. *Environmental Engineering Management* 13, 3083-3096.

Huang, M., Han, W., Wan, J., Ma, Y., Chen, X., 2016. Multi-objective optimisation for design and operation of anaerobic digestion using GA-ANN and NSGA-II. *Journal of Chemical. Technol. Biotechnology.* 91 (1), 226–233.

Ilangovan, P., Begum, M. S., & Srividhya, P. K. (2023). Development of online monitoring devices and performance evaluation of biogas plants using enhanced methane prediction algorithm (EMPA). *Sustainable Energy Technologies and Assessments*, *56*, 103041.

Ilo OP, Simatele MD, Nkomo SPL, Mkhize NM, Prabhu NG. The benefits of water hyacinth (eichhornia crassipes) for Southern Africa: a review. *Sustainability* 2020; 12:9222.

International Energy Agency, IEA Bioenergy Task 37 Country Reports Summary 2015, 2016. http://www.ieabioenergy.com/wp-content/uploads/2015/ 01/IEA-Bioenergy-Task-37-Country-Report-Summary-2014_Final.pdf.

International Renewable Energy Agency (IRENA) Renewable Energy Statistics 2023.

Jayaraj, S., Deepanraj, B., & Sivasubramanian, V. (2014, May). Study on the effect of pH on biogas production from food waste by anaerobic digestion. In *Proceedings of the 9th Annual Green Energy Conference, Tianjin, China* (pp. 25-28).

Jeong Y, Hermanowicz SW, Park C. Treatment of food waste recycling wastewater using anaerobic ceramic membrane bioreactor for biogas production in mainstream treatment process of domestic wastewater. *Water Resources* 2017; 123:86–95.

Jeong, K., Abbas, A., Shin, J., Son, M., Kim, Y.M., Cho, K.H., 2021. Prediction of biogas production in anaerobic co-digestion of organic wastes using deep learning models. *Water Resources.* 205, 117697.

Jia, R., Song, Y.C., Piao, D.M., Kim, K., Lee, C.Y., Park, J., 2022. Exploration of deep learning models for real-time monitoring of state and performance of anaerobic digestion with online sensors. *Bioresource Technology*. 363, 127908

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202.

Jomova, K., Valko, M., 2011. Advances in metal-induced oxidative stress and human disease. Toxicology 283, 65–87.

Juˇzniˇc-Zonta, Z., ˇ Kocijan, J., Flotats, X., Vreˇcko, D., 2012. Multi-criteria analyses of wastewater treatment bioprocesses under an uncertainty and a multiplicity of steady states. *Water Resources*. 46 (18), 6121–6131

Kalogirou, A.S., Artificial intelligence for the modeling and control of combustion processes: a review, Progress in Energy and Combustion Science, Volume 29, Issue 6, 2003, Pages 515-566, ISSN 0360-1285,

Kana, E. G., Oloke, J. K., Lateef, A., & Adesiyan, M. O. (2012). Modeling and optimization of biogas production on saw dust and other co-substrates using artificial neural network and genetic algorithm. *Renewable energy*, *46*, 276-281.

Kapoor, R.M., Vijay, V.K., Evaluation of Existing Low-Cost Gas Bottling Systems for Vehicles Use Adaption in Developing Economies, 2013

Karagiannidis, A., Perkoulidis, G., 2009. A multi-criterion ranking of different technologies for the anaerobic digestion for energy recovery of the organic fraction of municipal solid wastes. *Bioresource. Technology*. 100, 2355–2360.

Kashyap, D.R., Dadhich, K.S., Sharma, S.K., 2003. Biomethanation under psychrophilic conditions: a review. *Bioresource. Technology*. 87, 147–153

Kaza, S., Yao, L., Bhada-Tata, P., & Van Woerden, F. (2018). *What a waste 2.0: a global snapshot of solid waste management to 2050*. World Bank Publications.

Kazemi.P., Steyer. J-P., Bengoa.C., Font.J., Giralt.J. Robust data-driven soft sensors for online monitoring of volatile fatty acids in anaerobic digestion processes. *Processes,* 8 (1) (2020b), p. 67

Kazemi.P., Bengoa C., Steyer. J-P., Giralt, J., Data-driven techniques for fault detection in anaerobic digestion process. *Process Safety and Environmental Protection*., 146 (2021), pp. 905-915

Khalid, A., Arshad, M., Anjum, M., Mahmood, T., Dawson, L. (2011). The Anaerobic Digestion of Solid Organic Waste. *Waste Management*, 1737-1744.

Khan, M., Chuenchart W., Surendra K., Khanal, S. (2023). Applications of artificial intelligence in anaerobic co-digestion: Recent advances and prospects. Bioresource Technology 370, p.128501

Kim, J.K., Oh, B.R., Chun, Y.N., Kim, S.W., 2006. Effects of temperature and hydraulic retention time on anaerobic digestion of food waste. *Journal of. Bioscience. Bioengineering*. 102, 328–332.

Kim, J., Kim, B. S., & Savarese, S. (2012, January). Comparing image classification methods: K-nearest-neighbour and support-vector-machines. In *Proceedings of the 6th WSEAS international conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics* (pp. 133-138).

Kumar Khanal, S., Lü, F., Wong, J.W.C., Wu, D., Oechsner, H., 2021. Anaerobic Digestion Beyond Biogas. *Bioresource. Technology*. 125378 https://doi.org/10.1016/j. biortech.2021.125378

Kumar M., Dutta S., You S., Luo G., Zhang S., Show P.L., Sawarkar A.D., Singh.L., Tsang.D.C.W., (2021). A critical review on biochar for enhancing biogas production from anaerobic digestion of food waste and sludge, *Journal of Cleaner Production*, Volume 305,2021,127143, ISSN 0959-6526.

Kunatsa, T., & Xia, X. (2022). A review on anaerobic digestion with focus on the role of biomass co-digestion, modelling and optimization on biogas production and enhancement. *Bioresource technology*, 344, 126311

Li, Y. F., Kwok, J., & Zhou, Z. H. (2010, July). Cost-sensitive semi-supervised support vector machine. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 24, No. 1, pp. 500-505).

Li, D., Liu, S., Mi, L., Li, Z., Yuan, Y., Yan, Z., & Liu, X. (2015). Effects of feedstock ratio and organic loading rate on the anaerobic mesophilic co-digestion of rice straw and cow manure. *Bioresource Technology*, *189*, 319-326.

Li, Z., & Yan, X. (2019). Ensemble model of wastewater treatment plant based on rich diversity of principal component determining by genetic algorithm for status monitoring. *Control Engineering Practice*, *88*, 38-51.

Li, J., Zhang, L., Li, C., Tian, H., Ning, J., Zhang, J., Tong, Y.W., Wang, X., 2022. Data-driven based in-depth interpretation and inverse design of anaerobic digestion for CH4-rich biogas production. ACS ES&T Eng. 2 (4), 642–652.

Logan, M., Visvanathan, C., 2019. Management strategies for anaerobic digestate of organic fraction of municipal solid waste: Current status and future prospects. *Waste Management. Resources.* 37, 27–39

Lohani, S., Dhungana, B., Horn, H., Khatiwada, D. (2021). Small-scale biogas technology and clean cooking fuel: assessing the potential and links with SDGs in low-income countries. A case study of Nepal. *Sustainable Energy Technology Assessment*, 46, 101301.

Long, F., Wang, L., Cai, W., Lesnik, K., Liu, H., 2021. Predicting the performance of anaerobic digestion using machine learning algorithms and genomic data, *Water Resources*., 199, 117182.

Masahiko, H., Cian, F., Lall, U. (2019). Leveraging global and local data sources for flood hazard assessment and mitigation: an application of machine learning to Manila. United Nations Office for Disaster Risk Reduction, Contributing Paper to GAR 2019

Mehryar, E., Ding, W., Hemmat, A., Hassan, M., Talha, Z., Kafashan, J., Huang, H., 2017. Modeling and multiresponse optimization for anaerobic codigestion of oil refinery wastewater and chicken manure by using artificial neural network and the Taguchi method, 2017 *BioMed Res. Int.* 1–15

Molina.F., Castellano, C., Garcia, E., Roca, J., Lema, M., 2009. Selection of variables for on-line monitoring, diagnosis, and control of anaerobic digestion processes. *Water Science and Technology* Volume 60, Issue 3

Mougari.N.E., Largeau. J.F., Himrane. N., Hachemi. M., Tazerout.,2021. Application of artificial neural network and kinetic modeling for the prediction of biogas and methane production in anaerobic digestion of several organic wastes. *International Journal of Green Energy*, Volume 18, Issue 15.

Moller, ¨ K., Müller, T., 2012. Effects of anaerobic digestion on digestate nutrient availability and crop growth: A review. Eng. Life Sci. 12, 242–257. https://doi.org/ 10.1002/elsc.201100085.

Mukasine, A., Sibomana, L., Jayavel, K., Nkurikiyeyezu, K., & Hitimana, E. (2024). Maximizing Biogas Yield Using an Optimized Stacking Ensemble Machine Learning Approach. *Energies*, *17*(2), 364.

Nair, V.V., Dhar, H., Kumar, S., Thalla, A.K., Mukherjee, S., Wong, J.W.C., 2016. Artificial neural network-based modeling to evaluate methane yield from biogas in a laboratory-scale anaerobic bioreactor. *Bioresource. Technology*. 217, 90–99.

Obaideen K., Abdelkareem A.M., Wilberforce.T., Elsaid.K, Taha S.E., Maghrabieg H.M., Olabi. A.G., (2022). Biogas role in the achievement of sustainable development goals: Evaluation, Challenges, and Guidelines. *Journal of Taiwan Institute of Chemical Engineers* 131 (2022) 104207

Offie I., Piadeh F., Behzadian K., Campos L., Yaman R. (2022). Real-Time monitoring of decentralized Anaerobic Digestion using Artificial Intelligence-based framework. *International Conference on Resource Sustainability (icRS 2022)*, pp. 1-4.

Offie, I., Piadeh, F., Behzadian, K., Campos, L., Yaman, R. (2023). Development of an artificial intelligence-based framework for biogas generation from a micro anaerobic digestion plant. *Journal of Waste Management,* 158, pp 66-75

Okoro OV, Sun Z. Desulphurization of biogas: a systematic qualitative and economic-based quantitative review of alternative strategies. *ChemEngineering* 2019; 3:76

Oloko-Oba, M.I., Taiwo, A.E., Ajala, S.O., Solomon, B.O., Betiku. E. Performance evaluation of three different-shaped bio-digesters for biogas production and optimization by artificial neural network integrated with genetic algorithm. *Sustainable Energy Technological. Assessments*, 26 (2018), pp. 116-124

Onu P., Mbohwa.C., and Pradhan A., "An analysis of the application of machine learning techniques in anaerobic digestion," 2023 *International Conference on Control, Automation and Diagnosis* (ICCAD), Rome, Italy, 2023, pp. 1-6,

Orhorhoro, K.E., Ebunilo, P.O., Sadjere, E.G., Experimental Determination of Effect of Total Solid (TS) and Volatile Solid (VS) on Biogas Yield *American Journal of Modern Energy* 2017; 3(6): 131-135.

Ozkaya, B., Demir, A., & Bilgili, M. S. (2007). Neural network prediction model for the methane fraction in biogas from field-scale landfill bioreactors. *Environmental Modelling & Software*, *22*(6), 815-822.

Palaniswamy, D., Ramesh, G., Sivasankaran, S., Kathiravan, N., 2017. Optimizing biogas from food waste using a neural network model. In: *Proceedings of the Institution of Civil Engineers-Municipal Engineer.* Thomas Telford Ltd, pp. 221–229.

Panuccio, M.R., Attin` a, E., Basile, C., Mallamaci, C., Muscolo, A., 2016. Use of Recalcitrant Agriculture Wastes to Produce Biogas and Feasible Biofertilizer. Waste Biomass Valori 7, 267–280. https://doi.org/10.1007/s12649-015-9445-5

Paolini V, Petracchini F, Segreto M, Tomassetti L, Naja N, Cecinato A. 2018. Environmental impact of biogas: a short review of current knowledge. *Journal of Environmental Science Health* Part A 2018; 53:899–906

Parikh, S. M., & Shah, M. K. (2020). Classification of sentiment analysis using machine learning. In *Innovative Data Communication Technologies and Application: ICIDCA 2019* (pp. 76-86). Springer International Publishing.

Park, C., Lee, C., Kim, S., Chen, Y., Chase, H.A., 2005. Upgrading of anaerobic digestion by incorporating two different hydrolysis processes. *Journal of Bioscience. Bioengineering*. 100, 164–167.

Park.J., Jun.H., Heo.T-Y.,.2021. Retraining prior state performances of anaerobic digestion improves prediction accuracy of methane yield in various machine learning models. *Applied. Energy*, 298 (2021), Article 117250

Parthiban, A., Sathish, S., Suthan, R., Sathish, T., Rajasimman, M., Vijayan, V., & Jayaprabakar, J. (2023). Modelling and optimization of thermophilic anaerobic digestion using biowaste. *Environmental Research*, *220*, 115075.

Pei, Z., Liu, S., Jing, Z., Zhang, Y., Wang, J., Liu, J., Wang, Y., Guo, W., Li, Y., Feng, L., Zhou, H., Li, G., Han, Y., Liu, D., Pan, J., 2022. Understanding of the interrelationship between methane production and microorganisms in high-solid anaerobic co-digestion using microbial analysis and machine learning. *Journal of. Cleaner Production*., 373 (2022), Article 133848

Pell Frischmann Consultants Ltd, 2012. Enhancement and treatment of digestates from anaerobic digestion, UK.

Pfeiffer, D., Thr¨ an, D., 2018. One Century of Bioenergy in Germany: Wildcard and Advanced Technology. *Chem Ing Tech* (Weinh) 90, 1676–1698. https://doi.org/ 10.1002/cite.201800154

Piadeh F., Behzadian K., Alani. A., A critical review of real-time modelling of flood forecasting in urban drainage systems. *Journal of Hydrology*. (2022), Article 127476

Pramanik, Sagor Kumar, Fatitah Binti Suja, Shahrom Md Zain, Biplob Kumar Primanik (2019). The anaerobic digestion process of biogas production from food waste: Prospects and constraints. *Bioresource Technology Rep*. 8.

Qdais.A.H., Hani B.K., Shatnawi.N., Modelling, and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm. *Resources, Conservation and Recycling*, 54 (2010), 359-363.

Ramachandran, A., Rustum, R., Adeloye. A.J. Review of anaerobic digestion modeling and optimization using nature-inspired techniques. *Processes,* 7 (2019), pp. 1-12

Rasapoor, M. Young.B. (2020). Recognizing the challenges of anaerobic digestion: Critical steps towards improving biogas production. *Fuel* 261

Rasi S, Veijanen A, Rintala J. Trace compounds of biogas from different biogas production plants. Energy 2007; 32:1375–80

Rego, A. S., Leite, S. A. F., Leite, B. S., Grillo, A. V., & Santosa, B. F. (2019). Artificial neural network modelling for biogas production in biodigesters. *Chemical Engineering*. Trans, 74.

Riau, V., Ángeles De la Rubia. M., Pérez. M., Temperature-phased anaerobic digestion (TPAD) to obtain class A biosolids: A semi-continuous study, *Bioresource Technology*, Volume 101, Issue 8, 2010, Pages 2706-2712,

Rico, C., Montes, J.A., Lobo, A., 2020. Dry batch anaerobic digestion of food waste in a box-type reactor system: inoculum preparation and reactor performance. *Journal of Cleaner Production.* 251, 119751.

Rigby, H., Smith, S.R., 2013. Nitrogen availability and indirect measurements of greenhouse gas emissions from aerobic and anaerobic biowaste digestates applied to agricultural soils. *Waste Management*. 33, 2641–2652. https://doi.org/10.1016/j. wasman.2013.08.005

Rincón, B., Borja, R., González, J. M., Portillo, M. C., & Sáiz-Jiménez, C. (2008). Influence of organic loading rate and hydraulic retention time on the performance, stability, and microbial communities of one-stage anaerobic digestion of two-phase olive mill solid residue. *Biochemical engineering journal*, *40*(2), 253-261.

Romero-Güiza, M. S., Vila, J., Mata-Alvarez, J., Chimenos, J. M., & Astals, S. (2016). The role of additives on anaerobic digestion: a review. *Renewable and Sustainable Energy Reviews*, *58*, 1486-1499.

Rongwong, W., Lee, J., Goh, K., Karahan, H. E., & Bae, T. H. (2018). Membrane-based technologies for post-treatment of anaerobic effluents. *NPJ Clean Water*, *1*(1), 21.

Ross, D., 2012. Understanding value of nutrient credits. BioCycle 53, 57. https://www. biocycle.net/understanding-value-of-nutrient-credits/.

Ruiz, G., Castellano, M., González, W., Roca, E., & Lema, J. M. (2004). Transient state detection and prediction of organic overload in anaerobic digestion process using statistical tools. *IFAC Proceedings Volumes*, *37*(3), 357-362

Saadatlu.A.E., Barzinpour. F., Yaghoubi.S., 2023 A sustainable municipal solid waste system under leachate treatment impact along with leakage control and source separation. *Process Safety and Environmental Protection,* Volume 169, January 2023, Pages 982-998

Sarsaiya, S., Jain, A., Awasthi, K.S., Duan, Y., Awasthi, K., Shi.J., Microbial dynamics for lignocellulosic waste bioconversion and its importance with modern circular economy, challenges, and future perspectives. *Bioresources. Technology*., 291 (2019), Article 121905

Sathish, S.; Vivekanandan, S. Parametric optimization for floating drum anaerobic bio-digester using Response Surface Methodology and Artificial Neural Network. *Alex. Eng. J.* 2016, 55, 3297–3307.

Sawatdeenarunat C, Surendra KC, Takara D, Oechsner H, Khanal SK. Anaerobic digestion of lignocellulosic biomass: challenges and opportunities. *Bioresource Technology* 2015; 178:178–86.

Scarlat N., Dallemand J-F., Monforti-Ferrario, F., Banja M., Motola V., Renewable energy policy framework and bioenergy contribution in the European union e an overview from national renewable energy action plans and progress reports, *Renewable. Sustainable. Energy Review*. 51 (2015) 969e985, https:// doi.org/10.1016/j.rser.2015.06.062

Scheftelowitz, M., Becker, R., Thr¨ an, D., 2018. Improved power provision from biomass: A retrospective on the impacts of German energy policy. *Biomass Bioenergy* 111, 1–12. https://doi.org/10.1016/j.biombioe.2018.01.010

Şenol. H., 2021. Methane yield prediction of ultrasonic pretreated sewage sludge by means of an artificial neural network. *Energy*, 215 (2021), Article 119173

Seo, K.W., Seo, J., Kim, K., Lim, S.J., Chung, J., 2021. Prediction of biogas production rate from dry anaerobic digestion of food waste: process-based approach vs. recurrent neural network black-box model. *Bioresource. Technology*. 341, 125829.

Shahsavar.M.M., Akrami.M., Gheibi M., Kavianpour B., Fathollahi-Fard.A.M., Behzadian K. Constructing a smart framework for supplying the biogas energy in green buildings using an integration of response surface methodology, artificial intelligence, and petri net modelling. *Energy Conversion and Management*. 248, (2021), 114794

Shahsavar, M.M., Akrami M., Kian Z., Gheibi, M., Fathollahi-Fard, A.M., Hajiaghaei-Keshteli, M., Behzadian, K. Bio-recovery of municipal plastic waste management based on an integrated decision-making framework *J. Ind. Eng. Chem*., 108 (2022), pp. 215-234.

Shaibur MR, Husain H, Arpon SH. Utilization of cow dung residues of biogas plant for sustainable development of a rural community. *Current Resources Environmental Sustainability* 2021; 3:100

Shargil, D., Gerstl, Z., Fine, P., Nitsan, I., Kurtzman, D., 2015. Impact of biosolids and wastewater effluent application to agricultural land on steroidal hormone content in lettuce plants. *Science. Total Environment*. 505, 357–366. https://doi.org/10.1016/j. scitotenv.2014.09.100

Sharma, A. K., Ghodke, P. K., Goyal, N., Nethaji, S., & Chen, W. H. (2022). Machine learning technology in biohydrogen production from agriculture waste: Recent advances and future perspectives. *Bioresource technology*, 128076.

Shehab, M., Abualigah, L., Jarrah, M.I., Alomari, O.A., Daoud, M.S., 2020. Artificial Intelligence in software engineering and inverse: Review. *Int. J. Comput. Integr*. Manuf. 33 (10–11), 1129–1144.

Silkina, A., Zacharof, M.-P., Hery, G., Nouvel, T., Lovitt, R.W., 2017. Formulation and utilisation of spent anaerobic digestate fluids for the growth and product formation of single cell algal cultures in heterotrophic and autotrophic conditions. *Bioresource. Technology*. 244, 1445–1455. https://doi.org/10.1016/j.biortech.2017.05.133

Soni A., Das P.K., Kumar P., 2022 A review on the municipal solid waste management status, challenges, and potential for the future Indian cities. *Environment, Development and Sustainability* (2022).

Sonwai, A., Pholchan, P., Tippayawong, N. (2023). Machine learning approach for determining and optimizing influential factors of biogas production from lignocellulosic biomass. Bioresource Technology, 383, p.129235.

Sorda, G., Sunak, Y., Madlener, R., 2013. An agent-based spatial simulation to evaluate the promotion of electricity from agricultural biogas plants in Germany. *Ecol. Econ*. 89, 43–60. https://doi.org/10.1016/j.ecolecon.2013.01.022.

St-Pierre, B., & Wright, A. D. G. (2014). Comparative metagenomic analysis of bacterial populations in three full-scale mesophilic anaerobic manure digesters. *Applied microbiology and biotechnology*, *98*, 2709-2717.

Studer I, Boeker C, Geist J. 2017. Physicochemical and microbiological indicators of surface water body contamination with different sources of digestate from biogas plants. *Ecology Indices* 2017; 77:314–22.

Strik, David P.B.T.B. Alexander M. Domnanovich, Loredana Zani, Rudolf Braun, Peter Holubar. Prediction of trace compounds in biogas from anaerobic digestion using the MATLAB Neural Network Toolbox, *Environmental Modelling & Software*, Volume 20, Issue 6, 2005, Pages 803-810, ISSN 1364-8152

Sun, M. T., Fan, X. L., Zhao, X. X., Fu, S. F., He, S., Manasa, M. R. K., & Guo, R. B. (2017). Effects of organic loading rate on biogas production from macroalgae: Performance and microbial community structure. *Bioresource technology*, *235*, 292-300

Taherzadeh MJ, Karimi K. Pretreatment of Lignocellulosic Wastes to Improve Ethanol and Biogas Production: A Review. *International Journal of Molecular Sciences*. 2008; 9(9):1621-1651. https://doi.org/10.3390/ijms9091621

Tang WZ. Physicochemical treatment of hazardous waste. CRC Press; 2003.

Tao X., Liu.H., Mao. C., Wang.C., Yap.J.B.H., Sepasgozar, S., Shirowzhan, S., Rose, T., Developing Shuffled Frog-Leaping Algorithm (SFLA) Method to Solve Power Load-Constrained TCRTO Problems in Civil Engineering. Advances in Civil Engineering Volume 2019, Article ID 1404636 https://doi.org/10.1155/2019/1404636

Tharwat, A. (2021), "Classification assessment methods*", Applied Computing and Informatics*, 17(1), pp.168-192.

Thran, ¨ D., Schaubach, K., Majer, S., Horschig, T., 2020. Governance of sustainability in the German biogas sector—adaptive management of the Renewable Energy Act between agriculture and the energy sector. Energy Sustain. Soc. 10, 3. https://doi. org/10.1186/s13705-019-0227-y

Tozlu, A., Emrah.Ö., Ayşegül A. 2016. Waste to energy technologies for municipal solid waste management in Gaziantep Alperen. *Renewable and Sustainable Energy Reviews*. Volume 54, 2016, Pages 809-815.

Trucchia, A., Frunzo, L., 2021. Surrogate based global sensitivity analysis of ADM1-based anaerobic digestion model. *Journal of. Environmental. Management*. 282, 111456

Tufaner, F., Avsar, Y., Gonüllü, M.T., 2017. Modeling of biogas production from cattle manure with co-digestion of different organic wastes using an artificial neural network. *Clean Technology. Environment*. Policy 19, 2255–2264

Tufaner, F., Demirci, Y., 2020.Prediction of biogas production rate from anaerobic hybrid reactor by artificial neural network and nonlinear regressions models. *Cleaner Technology Environment*. Policy, 22 (2020), pp. 713-724

Uddin, M. & Wright, M. (2023). Anaerobic digestion fundamentals, challenges, and technological advances. *Physical Sciences Reviews*, *8*(9), 2819-2837.

Ullah Khan I, Hafiz. D.O.M, Hashim H, Matsuura T, Ismail AF, Rezaei D. M, Wan Azelee I.2017. Biogas as a renewable energy fuel; A review of biogas upgrading, utilization and storage. *Energy Conversion Management* 2017; 150:277–94.

Ullah, M., Han, K., Hadi, F., Xu, J., Song, J., & Yu, D. J. (2021). PScL-HDeep: image-based prediction of protein subcellular location in human tissue using ensemble learning of handcrafted and deep learned features with two-layer feature selection. *Briefings in Bioinformatics*, *22*(6), bbab278.

UK Environment Agency, 2014. Quality Protocol. Anaerobic digestate. End of waste criteria for the production and use of quality outputs from anaerobic digestion of source-segregated biodegradable waste. UK Environment Agency, UK.

USDA, 2015. Announces Funding for Renewable Energy and Energy Efficiency Projects. https://www.rd.usda.gov/newsroom/news-release/usda-announces-funding-re newable-energy-and-efficiency-projects-6 (accessed 24 January 2024)

USDA, 2020. Rural energy America program renewable energy systems energy efficiency. accessed 30 January 2020. https://www.rd.usda.gov/programs-services/rural-energy.

US EPA, 2020a. AgSTAR Project Development Handbook. United States Environmental Protection Agency (EPA).

US EPA, 2020b. AgSTAR: Biogas Recovery in the Agriculture Sector. https://www.epa. gov/agstar (accessed 18 January 2024).

Epa, U.S., 2020c. Renewable Identification Numbers (RINs) under the Renewable Fuel Standard Program. https://www.epa.gov/renewable-fuel-standard-program/renewa ble-identification-numbers-rins-under-renewable-fuel-standard (accessed 25 January 2024).

US EPA, 2020d. Renewable Energy Certificates (RECs). https://19january2021snapshot. epa.gov/greenpower/renewable-energy-certificates-recs_.html (accessed 26 January 2024).

Usmani Z., Sharma M., Karpichev Y., Pandey A., Kuhad R.C., Bhat R., Punia R., Aghbashlo M., Tabatabaei M., Gupta V.K (2020). Advancement in valorisation technologies to improve utilization of bio-based waste in bioeconomy context, *Renewable and Sustainable Energy Reviews*, Volume 131,2020,109965, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2020.109965.

Tamburini E, Gaglio M, Castaldelli G, Fano EA. 2020. Biogas from agri-food and agricultural waste can appreciate agro-ecosystem services: the case study of Emilia Romagna region. *Sustainability* 2020; 12:8392.

Velivela A, Barham H, Bauer J, Roschke J, Daim TU, Meissner D. 2020. Biogas: converting waste to Energy. In: Daim TU, Meissner D, editors. Innovation management in the intelligent world: cases and tools. Cham: *Springer International Publishing*; 2020. p. 285–98.

Vembandasamy, K., Sasipriya, R., & Deepa, E. (2015). Heart diseases detection using Naive Bayes algorithm. *International Journal of Innovative Science, Engineering & Technology*, *2*(9), 441-444.

Verhoog R, Ghorbani A, Dijkema GP. 2016. Modelling socio-ecological systems with MAIA: a biogas infrastructure simulation. *Environmental Modelling Software* 2016; 81:72–85.

Wahba.S, Kaza.S, Ionkova. .K.M 2019 World Bank Blogs https://blogs.worldbank.org/sustainablecities/new-phenomenon-realizing-economic-growthwhile-cutting-waste-how

Wainaina S., Awasthi.K.M., Surrendra.S., Chen.H., Singh E., Kumar. A., Ravindran B., Awasthi.S.K., Liu.T., Duan. Y., Kumar.S., Zhang.Z., Taherzadeh.M.J., 2020. Resource recovery and circular economy from organic solid waste using aerobic and anaerobic digestion technologies. *Bioresource Technology* 2020, Volume 301, 122778

Walker M., H. Theaker, R. Yaman, D. Poggio, W. Nimmo, A. Bywater, M. Pourkashanian.2017. Assessment of Micro-Scale Anaerobic Digestion for Management of Urban Organic Waste: A Case Study in London, UK. *Journal of Waste Management*., 2017 (01) (2017), p. 036

Wang L., Long F., Wei.L., Liu.H. Prediction of anaerobic digestion performance and identification of critical operational parameters using machine learning algorithms. *Bioresource Technology* (2020), Volume 298, 122495

Wang, P., Peng, H., Adhikari, S., Higgins, B., Roy, P., Dai, W., & Shi, X. (2020). Enhancement of biogas production from wastewater sludge via anaerobic digestion assisted with biochar amendment. *Bioresource technology*, *309*, 123368.

Wang Y., Huntington T., Scown.D. C. Tree-Based Automated Machine Learning to Predict Biogas Production for Anaerobic Co-digestion of Organic Waste*. ACS Sustainable Chemistry & Engineering* 2021 9 (38), 12990-13000 DOI: 10.1021/acssuschemeng.1c04612

Wang, M., Liu, Y., Jiang, X., Fang, J., Lyu, Q., Wang, X., Yan, Z., 2021a. multi-omics reveal the structure and function of microbial community in co-digestion of corn straw and pig manure. J. Clean. Prod. 322, 129150.

Wang, Z., Peng, X., Xia, A. Shah, A., Huang, Y., Zhu, X., Zhu, X., Liao, Q. (2022). The role of machine learning to boost the bioenergy and biofuels conversion. *Bioresource Technology*, 343, p.126099

Wang, Z., Peng, X., Xia, A., Shah, A. A., Yan, H., Huang, Y., & Liao, Q. (2023). Comparison of machine learning methods for predicting the methane production from anaerobic digestion of lignocellulosic biomass. *Energy*, *263*, 125883.

Ward, A.J., Hobbs, P.J., Holliman, P.J., Jones, D.L., 2008. Optimization of the anaerobic digestion of agricultural resources. *Bioresource. Technology*. 99, 7928–7940

Weiland, P. (2000). Anaerobic waste digestion in Germany–Status and recent developments. *Biodegradation*, *11*, 415-421.

Winther T. (2008) The impact of electricity: development, desires, and dilemmas. Berghahn Books; 200

Wolf, C., Gaida, D., Stuhlsatz, A., McLoone, S., Bongards, M. (2010). Organic Acid Prediction in Biogas Plants Using UV/vis Spectroscopic Online-Measurements. In: Li, K., Li, X., Ma, S., Irwin, G.W. (eds) Life System Modeling and Intelligent Computing. ICSEE LSMS 2010 2010. *Communications in Computer and Information Science*, vol 97. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-15853-7_25

World Biogas Association. (2018). The contribution of Anaerobic Digestion and Biogas towards achieving the UN Sustainable Development Goals. Retrieved from www.worldbiogasassociation.org.

World Biogas Association. (2021). Biogas: Pathways to 2030. Retrieved from www.worldbiogasassociation.org.

Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand & Dan Steinberg (2007). Top 10 algorithms in data mining

Xu, W., Long, F., Zhao, H., Zhang, Y., Liang, D., Wang, L., Larson K., Cao, L., Zhang, Y., Liu, H. (2021). Performance prediction of ZVI-based anaerobic digestion reactor using machine learning algorithms. Waste Management, 121, pp. 59-66.

Xu, J., Yuan, X., Zhang, Y., Yu, S., Pang, Y., Zhang, T., ... & Ding, X. (2022). Real-time prediction of drilling forces inside lunar regolith based on recurrent neural networks. *Acta Astronautica*, *201*, 259-273.

Xu, R., Cao, J., Ye, T., Wang, S., Luo, J., Ni, B., Fang, F. (2022). Automated machine learning-based prediction of microplastics induced impacts on methane production in anaerobic digestion, Water Research, 223, p.118975.

Xue, S., Qiu, L., Guo, X., & Yao, Y. (2020). Effect of liquid digestate recirculation on biogas production and enzyme activities for anaerobic digestion of corn straw. *Water Science and Technology*, *82*(1), 144-156.

Yang, C. C., Prasher, S. O., Enright, P., Madramootoo, C., Burgess, M., Goel, P. K., & Callum, I. (2003). Application of decision tree technology for image classification using remote sensing data. *Agricultural Systems*, *76*(3), 1101-1117

Yang G., Zuang G., Zhuan R., Yang A., Wang Y., A critical review on sludge anaerobic digestion progress from 2012 to 2015 *Current Organic Chemistry*, Volume 20, Number 26, 2016, pp. 2762-2779(18)

Yang W., Li S., Qv. M., Dai. D, Liu.D, Wang W., Tang C., Zhu. L., Microalgal cultivation for the upgraded biogas by removing CO2, coupled with the treatment of slurry from anaerobic digestion: A review. *Bioresource Technology*., 364 (2022), Article 128118

Yasar A, Nazir S, Tabinda AB, Nazar M, Rasheed R, Afzaal M. Socio-economic, health and agri-culture benefits of rural household biogas plants in energy scarce developing countries: a case study from Pakistan. *Renewable Energy* 2017; 108:19– 25.

Yetilmezsoy, K., Ozkaya, B., & Cakmakci, M. (2011). Artificial intelligence-based prediction mod-els for environmental engineering. *Neural Network World*, *21*(3).

Yeltimezsoy K., Turkdogan F.I., Temizel I., Gunay.A., Development of Ann-Based Models to Pre-dict Biogas and Methane Productions in Anaerobic Treatment of Molasses Wastewater. *Interna-tional Journal of Green Energy*, Volume 10, 2013- Issue 9.

Yi-Fan, H., Chang-Zhu, Y., Jin-Feng, D., Wen-Hong, P., Jia-Kuang, Y., 2017. Modeling of ex-panded granular sludge bed reactor using artificial neural network. *Journal of. Environmental. Chemical. Engineering*. 5 (3), 2142–2150

Yildirim, O., Ozkaya, B., (2023). Prediction of biogas production of industrial scale anaerobic di-gestion plant by machine learning algorithms. *Chemosphere,* Volume 335,2023,138976, ISSN 0045-6535

Yüksel. N., Börklü.H.R., Sezer.H.K., Canyurt. O.E., Review of artificial intelligence applications in engineering design perspective. *Engineering Applications of Artificial Intelligence* 118 (2023), 105697

Zacharof, M.-P., Lovitt, R.W., 2014. The filtration characteristics of anaerobic digester effluents employing cross flow ceramic membrane microfiltration for nutrient recovery. *Desalination* 341, 27–37. https://doi.org/10.1016/j.desal.2014.02.034.

Zaied, B., Rashid, M., Nasrullah, M., Bari, B.S., Zularisam, A., Singh, L., Kumar, D., Krishnan, S., 2020. Prediction and optimization of biogas production from POME co digestion in solar bioreactor using artificial neural network coupled with particle swarm optimization (ANN-PSO). *Biomass Conversion. Biorefinery*. 1–16.

Zareei, S., & Khodaei, J. (2017). Modeling and optimization of biogas production from cow manure and maize straw using an adaptive neuro-fuzzy inference system. *Renewable Energy*, *114*, 423-427.

Zeinolabedini, M., Pazoki, M. & Saeid, P. Optimization of parameters that affect biogas production of anaerobic digestion using a particle swarm algorithm. *Int J Energy Water Res* **7**, 29–41 (2023). https://doi.org/10.1007/s42108-022-00226-9

Zeng J, Xu R, Sun R, Niu L, Liu Y, Zhou Y, Zeng W, Yue Z. Evaluation of methane emission flux from a typical biogas fermentation ecosystem in China. *Journal of Cleaner Production,* 2020; 257:120441

Zhang, C., Zhu, X., et al. GBKII: an imputation method for missing values. PAKDD (2007), pp. 1080-1087

Zhang, D., Yuntian, C. H. E. N., & Jin, M. E. N. G. (2018). Synthetic well logs generation via recurrent neural networks. Petroleum Exploration and Development, 45(4), 629-639.

Zhang, H., Wang, Z., & Liu, D. (2007). Robust exponential stability of recurrent neural networks with multiple time-varying delays. *IEEE Transactions on Circuits and Systems II: Express Briefs*, *54*(8), 730-734

Zhang.L., Loh K., Zhang.J. Enhanced biogas production from anaerobic digestion of solid organic wastes: Current status and prospects Bioresource. Technology. Reports, 5 (2019), pp. 280-296.

Zhang, S., Z. Jin, X. Zhu. Missing data imputation by utilizing information within incomplete instances. J. Syst. Software, 84 (3) (2011), pp. 452-459

Zhang, S., Li, M., Zong, X., Zhu, X., Cheng.D., Learning k for kNN classification ACM Trans. Intelligent. System. Technology., 8 (3) (2017), p. 43

Zhang, Y., Feng, Y., Ren, Z., Zuo, R., Zhang, T., Li, Y., & Pan, J. (2023). Tree-based machine learning model for visualizing complex relationships between biochar properties and anaerobic digestion. Bioresource Technology, 374, 128746

7 Appendices

Table A1. Optimal cyclic pattern of feeding to the micro-AD plant and waste composition to obtain maximum biogas generation in a four-day cycle.

Parameter	Days							
	Predictors (input data) between t-3 and t				Prediction of t+1			
	t-3	t-2	t-1	t	t+1			
Feed	0	0	0	80				
Biogas	-	-	4.11	4.08	4.52			
Catering	-	-	55	5				
Oat	-	-	20	0				
Liner	-	-	0	0				
Water	0	0	15	0				
Parameter	Days							
	Predictors (input data) between t-2 and t+1				Prediction of t+2			
		t-2	t-1	t	t+1	t+2		
Feed		0	0	80	0			
Biogas		-	4.11	4.08	4.52	4.23		
Catering		-	55	5	0			
Oat		-	20	0	0			
Liner		-	0	0	0			
Water		0	15	0	0			
Parameter	Days							
	Predictors (input data) between t-1 and t+2				Prediction of t+3			
			t-1	t	t+1	t+2	t+3	
Feed			0	80	0	0		
Biogas			4.11	4.08	4.52	4.23	4.11	
Catering			55	5	0	0		
Oat			20	0	0	0		
Liner			0	0	0	0		
Water			15	0	0	0		

Parameter	Days							
	Predictors (input data) between t and t+3				Prediction for t+4			
				t	t+1	t+2	t+3	t+4
Feed				80	0	0	0	
Biogas				4.08	4.52	4.23	4.11	4.08
Catering				5	0	0	55	
Oat				0	0	0	20	
Liner				0	0	0	0	
Water				0	0	0	15	

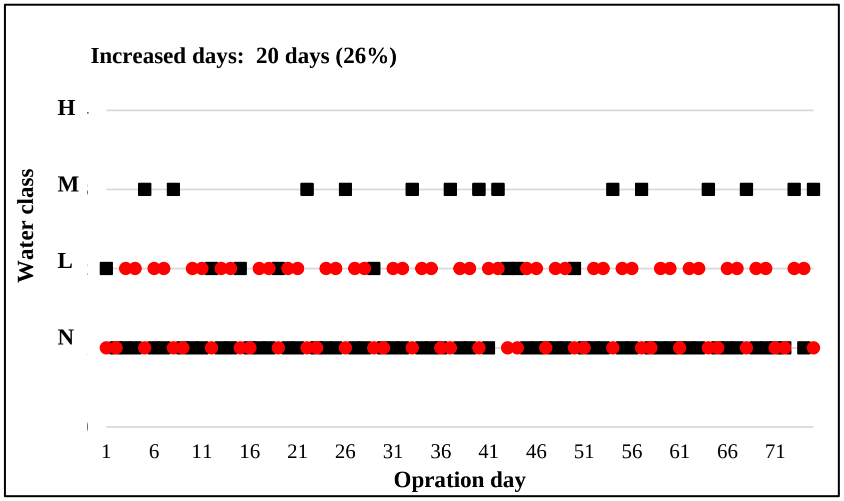
Table A2. Selected weak learning data mining models.

Selected methods	Description	Optimised hyperparameters
DA	Multiple linear regression expressing one dependent variable as a combination of other features or measurements	- Delta: Linear coefficient threshold - Gamma: Amount of regularisation
DT	The regression tree utilised a top-down recursive tree of an inner node. The decision tree model is divided into smaller subgroups until ultimately separated into an exclusive mutual subset.	- Minimum leaf size: Minimum number of leaf node observations
GPR	The kriging method providing the best linear unbiased prediction at unsampled locations	- Sigma: Initial value for the noise standard deviation
KNN	Non-parametric method finding the closest neighbourhoods based on similarity	- Distance: Neighbour search method - Neighbours number: Number of nearest neighbours in observant data to find for classifying each point when predicting
NB	Supervised learning method applying the theory of Bayes with strong independence assumptions between the different features	- Kernel distribution: Approach of data distribution and data smoothing - Width: Regulating width of Kernel smoothing window
SVM	Linear classification by splitting the data into subsets, e.g., pattern recognition and data classification based on the statistical learning theory and structural risk minimisation principle	- Kernel scale: Approach of data distribution and data smoothing - Box constraint: controller of the maximum penalty aiding to prevent overfitting

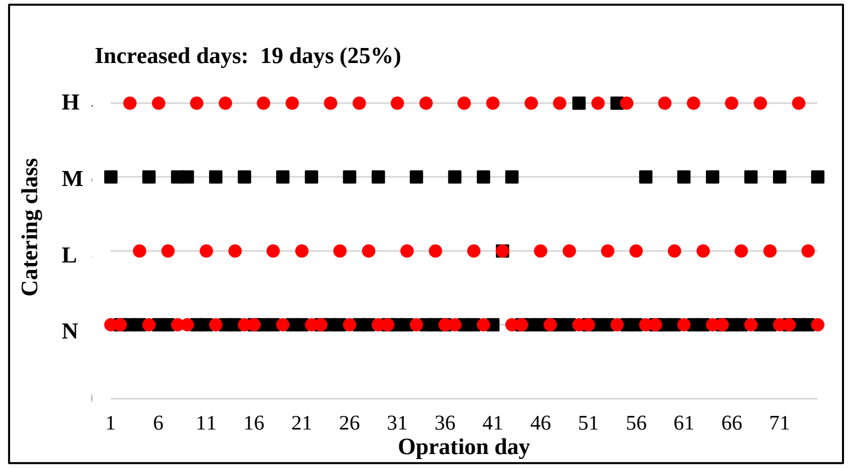
DA: Discriminant Analysis; DT: Decision Tree; GPR: Gaussian Process Regression; KNN: K-Nearest Neighbourhood; NB: Native Bayes
SVM: Supervised Vector Machine with Error-Correcting Output

Table A3. Ratio of different used biogas loads used for different steps of model development.

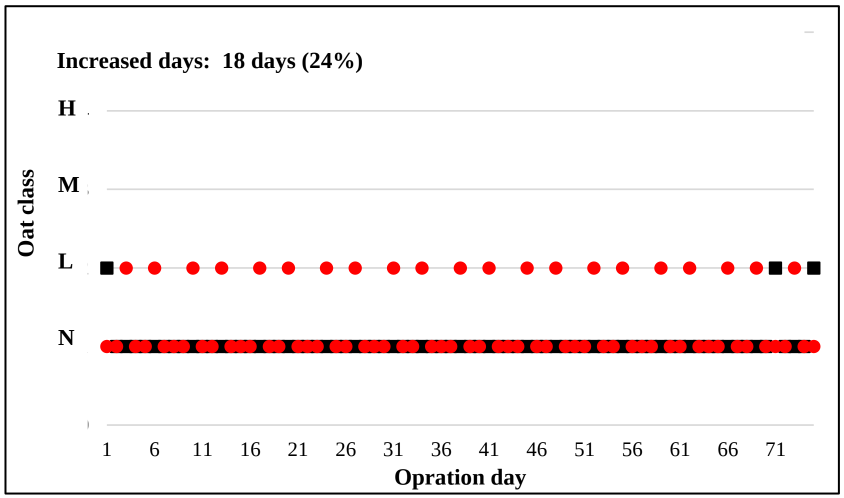
Step	Yielded biogas class (%)		
	Low	Medium	High
WLDM training	24.34	46.71	28.95
WLDM testing	26.32	34.21	39.47
Ensemble testing	22.37	46.05	31.58



(a)



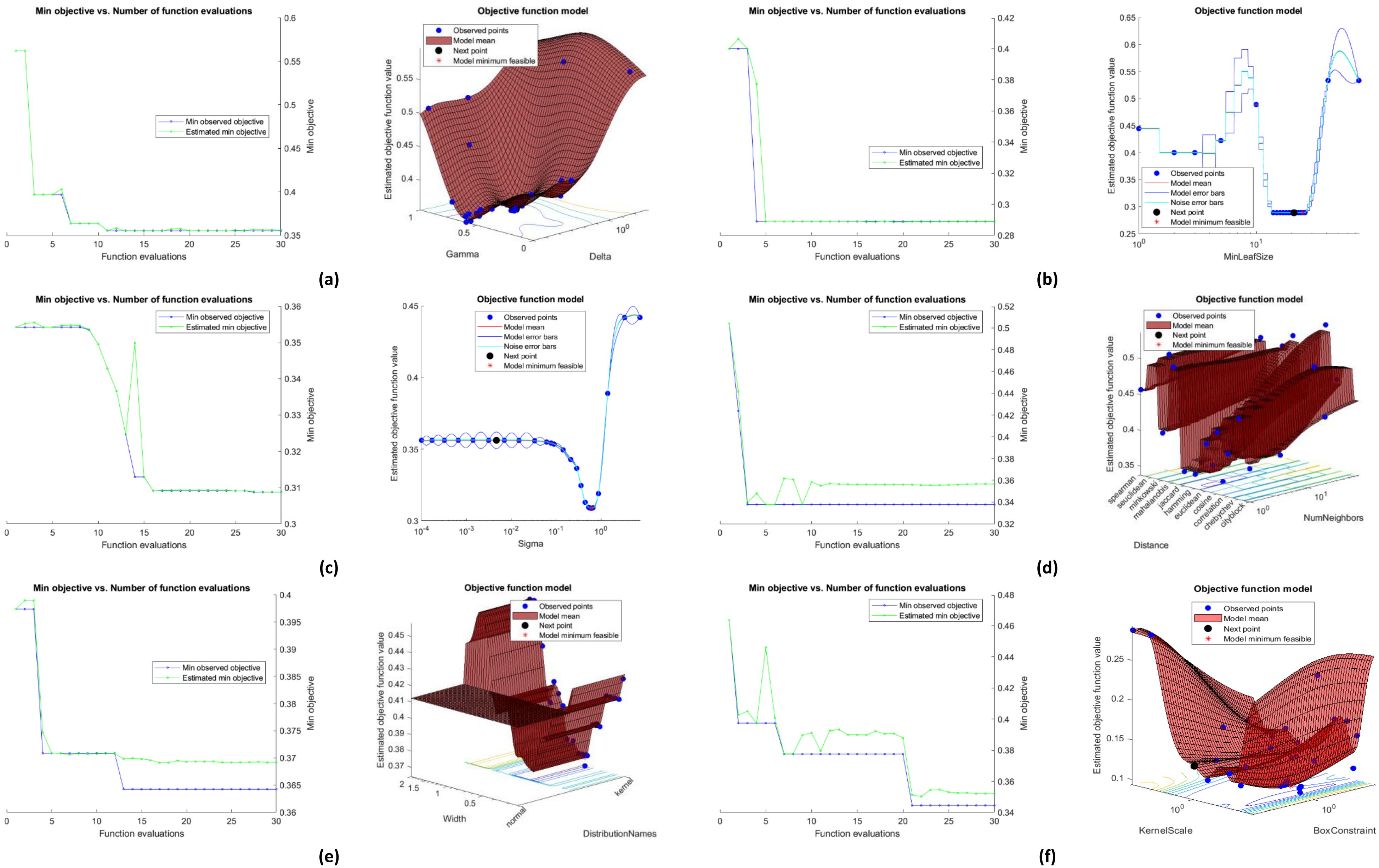
(b)



(c)

■ Measured Data ● Predicted Data

Figure A2. Comparison between optimal weekly operation pattern and testing event (76 days): (a) adding water, (b) adding catering material, (c) adding oat material.



Min objective vs. Number of function evaluations
Min observed objective
Estimated min objective
Function evaluations
Min objective
Objective function model
Observed points
Model mean
Next point
Model minimum feasible
Estimated objective function value
Gamma
Delta
(a)
Min objective vs. Number of function evaluations
Min observed objective
Estimated min objective
Function evaluations
Min objective
Objective function model
Observed points
Model mean
Model error bars
Noise error bars
Next point
Model minimum feasible
Estimated objective function value
MinLeafSize
(b)
Min objective vs. Number of function evaluations
Min observed objective
Estimated min objective
Function evaluations
Min objective
Objective function model
Observed points
Model mean
Model error bars
Noise error bars
Next point
Model minimum feasible
Estimated objective function value
Sigma
(c)
Min objective vs. Number of function evaluations
Min observed objective
Estimated min objective
Function evaluations
Min objective
Objective function model
Observed points
Model mean
Next point
Model minimum feasible
Estimated objective function value
spearman
seuclidean
minkowski
mahalanobis
jaccard
hamming
euclidean
cosine
correlation
chebychev
cityblock
Distance
NumNeighbors
(d)
Min objective vs. Number of function evaluations
Min observed objective
Estimated min objective
Function evaluations
Min objective
Objective function model
Observed points
Model mean
Next point
Model minimum feasible
Estimated objective function value
normal
kernel
Width
DistributionNames
(e)
Min objective vs. Number of function evaluations
Min observed objective
Estimated min objective
Function evaluations
Min objective
Objective function model
Observed points
Model mean
Next point
Model minimum feasible
Estimated objective function value
KernelScale
BoxConstraint
(f)

Figure A3. Optimisation results of the WLDM models for: (a) discriminant analysis, (b) decision tree, (c) Gaussian process regression, (d) k-nearest neighbourhood, (e) naive bayes, (f) supervised vector machine