Real-time monitoring of decentralised Anaerobic Digestion using Artificial Intelligence-based framework

Real-time Monitoring of Decentralised Anaerobic Digestion using Artificial Intelligence-Based Framework

Ikechukwu Offie[1], Farzad Piadeh[1], Kourosh Behzadian[1], Amir Alani[1], Rokiah Yaman[2], Luiza Campus[3]

[1] School of Computing and Engineering, University of West London, Ealing, London, W5 5RF, UK

[2] Leap AD Micro Company, London, UK

[3] Civil, Environmental and Geomatic Engineering, University College London, Gower St, London WC1E6BT, UK

Abstract: This paper presents an Artificial Intelligence (AI)-based framework for real-time monitoring and improving the operation of an Anaerobic Digestion (AD) system in producing biogas. This was achieved using historic data obtained from a decentralised AD plant located in Camley-Central London to develop a recurrent neural network (RNN) model based on AI to predict biogas production with respect to lag time. The dataset obtained from the AD plant had a wide range of missing values, which hindered the accurate prediction of biogas. This study evaluates different data mining techniques for infilling missing data. The Recurrent Neural Network (RNN) Model was then developed for predicting biogas with respect to various lag times. The results show both Kriging and Linear Regression techniques have the best performance, and they were used to infill the missing data. The results also show biogas production can be accurately predicted in real-time operation using a NARX model based on the feed data including organic food composition such as oats, soaked liners, catering and water added.

Keywords: Anaerobic Digestion, Biogas prediction, Neural Network Based State Estimation, Organic Waste, Recurrent Neural Network, Root Mean Square Error.

1. Introduction

Organic Waste is a major source of concern in the world today following the rapid population growth, economic development and urbanisation which has continuously occurred in different parts of the world over the years (Arun and Sivashanmugam., 2017). It accounts mainly for approximately 105 billion tonnes of the total Municipal Solid Waste generated on an annual basis globally (World Biogas Association, 2021). Improper and Inefficient waste treatment can lead to a series of environmental problems such as environmental pollution, ecosystem destruction, harm to human health and depletion of natural resources (Triassi et al., 2015; Laurent et al., 2014). The poor management of organic wastes also has the potential to contribute to climate change through the emission of greenhouse gases (GHGs) into the atmosphere (CIWEM., 2021). The effect of this has compelled nations and governments to invest more financial and material resources for the remediation of organic wastes in recent years (Wainaina et al., 2020).

Currently, serious efforts are being made to revolutionise the waste management industry towards achieving sustainability and profitability using advanced technologies and recycling methods such as Anaerobic Digestion, Composting and Incineration amongst others as they have been identified as better alternatives to the landfill system (Wainaina et al., 2020). Intelligent systems have also been incorporated into these technologies to improve waste management (Guo et al., 2021). The Introduction of Intelligent System into these technologies is not only conducive for environmental protection and sustainable development but is also critical to forging a circular economy (Felix et al., 2019). In a recent study conducted by Wainaina et al in 2020, the Anaerobic Digestion (AD) technology was regarded as an established biological processing technique suitable for stabilising a plethora of organic solid wastes coupled with the recovery of energy and nutrients attractive. It has also been identified as one of the foremost cost-effective biological treatments of organic wastes serving as an alternative to the landfill system of managing organic wastes (Wainaina et al., 2020). This is mainly due to its ability to generate nutrient rich digestate useful in improving soil fertility whereas lessening natural impacts of the waste transfer (Wainaina et al., 2020). The AD technology has also been revealed to give room for vitality recuperation (Wainaina et al., 2020).

Similarly, the composting technology has also been identified as a highly effective technique suitable for the management of organic wastes (Shyamala, Belagali., 2012). The effectiveness of this technique is based on its ability to treat lumps of organic wastes through biological action, converting them into stable substances such as humus and digestate applied to the soil as fertilizer to improve the fertility of the soil (Wainaina et al., 2020). Through this means, million tonnes of Nitrogen and Organic Carbon will be saved (Luis et al., 2019). This was revealed in statistical studies conducted by Luis et al. (2019) which indicated that most of these nutrients are lost through landfilling organic waste as only 5% of Organic Solid Wastes are currently being recycled across the European Countries. Further studies by Luis et al. (2019) also indicated that the recycling of more organic wastes using the composting technology could approximately replace 30% of chemical fertilizer added to the soil, which is equal to 1.8 million tonnes of phosphate fertilizer every year.

The incineration technology has also received worldwide attention having been observed to be a better alternative to the landfill technology (RecyclingInside, 2021). This is also due to its ability to reduce organic wastes, converting them into energy which can be used for several purposes beneficial to man (RecyclingInside, 2021)**.** The incineration technology has also been useful especially in the elimination of groundwater contamination through the diversion of organic wastes from landfills (Wright, 2020)**.** It has mostly been applied in developed countries where it is economically feasible for treating wastes, reducing, and burning them into less harmful substances **(**Wainaina et al., 2020).

Following the observations made on each of these technologies, it can be deduced that these technologies have been extensively applied in the management of organic wastes and have proven to be successful. However, the Anaerobic Digestion (AD) technology has been identified by Bahreini et al. (2020) as one of the most important waste management approaches that removes wastes from the environment and converts them into biogas. It is also capable of delivering both de-fossilisation and decarbonisation through the conversion of wastes to renewable energy thereby reducing the need for fossil fuel utilisation (World Biogas Association., 2021). This also creates opportunities for carbon to be recycled in various forms to meet the various needs of man and UN sustainable development goals (World Biogas Association., 2021).

The multi-faceted nature of the AD technology has rendered it a highly ranked method in the waste management industry globally and an excellent tool for the realisation of circular economy (World Biogas Association., 2021). On the other hand, the AD technology has been observed to have some technical limitations which affects the efficient production of biogas (Park et al., 2005). These limitations include long retention time and low removal efficiency of organic compounds (Park et al.,2005) Following these limitations, the need for the introduction of automated systems like Artificial Intelligence models into the AD system is of great significance in improving the efficiency of the AD system as they have been observed to be better alternatives to the mechanistic models especially for biogas energy generation (Shahsavar et al. 2021). This is due to their ability to give a high prediction accuracy when applied to complex non-linear problems within a short time unlike the mechanistic models which have been used in the past to address the limitations of the AD but were observed to be complex, analytically insolvable, time consuming, and not useful in controlling AD systems.

Hence, this study focuses mainly on the development of a smart framework for real-time monitoring and operation of AD system and improving the operation of the AD system in producing a higher volume of biogas within a reduced timeframe by using a Non-Linear Autoregressive Artificial Neural Network (NARX-ANN) Model. To achieve this, a decentralised AD plant in Camley-Central London was used as the pilot study where historic data was obtained from the decentralised AD plant and used to develop the ANN Model. However, the data obtained from the AD plant was observed to have a wide range of missing data which hindered the accurate prediction of biogas. This research study however, explored different techniques for infilling missing data to obtain the best range of data suitable for the development of the ANN Model. The paper is organised as follows; In Section 2, the different techniques adopted for infilling missing data will be presented and described. The type of ANN Model developed for monitoring and improving the efficiency of the AD System will be also presented in this section. In Section 3, the results obtained from the model development and testing will be presented and discussed in detail. Finally, the conclusion of the paper will be presented in Section 4.

2. Methodology

Artificial Neural Networks (ANNs) and different Data Mining techniques were employed and used as the core tools to prepare the data for the prediction of generated biogas. This was done on the MATLAB platform which provides the different functions for prediction purposes. The employment of these core tools was carried out after the collection of data from the AD plant and the selection of relevant data relevant for the development of the model as presented in the flowchart in Figure 1. The different data mining techniques were tested on to determine

the most appropriate technique suitable for infilling the missing data. For infilling data, the whole data was first divided into the two classes namely: data with feeding inclusive and data without feeding. The weak learner data mining techniques were used to model the relationship between feeding and generated biogas of which the best model to infill the missing data (data with feeding but no biogas). After this was done, other datapoints (data where feeding is zero and biogas is not read) were infilled based on linear regression of the remaining total biogas data read. The data mining techniques used for infilling missing data include Random Forest, K-Nearest Neighbor, Support Vector Machine, Naïve Bayes, Kriging, Feed Forward Neural Network and Linear Regression. This is shown in Figure 3 under the Results and Discussions Section. The Kriging technique was however selected and used to infill the missing biogas values in the data as it demonstrated to have the least range of fluctuations in its results compared to other data mining techniques.
Following this, a time-series Recurrent Neural Network (RNN) also known as Non-Linear Autoregressive Neural Network (NARX) Model was developed on the MATLAB platform. The development of this model was achieved taking into consideration, the influencing parameters of feed, composition, water, and biogas where the feed, oats composition, catering composition, liner-soaked composition and water added. These influencing parameters otherwise known as the Input parameters were trained as a regressor to predict the volume of biogas generated with respect to various lag times. The structure of the NARX Model can be seen in Figure 2 below. The NARX Model was used to specify the lag time for each of the input variables.

The model parameters were optimised afterwards using shuffled frog leaping algorithm (SFLA) to optimise the best lag times. This is a memetic and nature-based algorithm with ability to search in both local and global search space where each lag time represents one frog (Bui et al., 2020). Thus, 4 samples were used for exploration and 4 samples were used for exploitation. The performance of the model was evaluated using two indicators. These two indicators include Root Mean Square Error (RMSE) and Neural Network Based State Estimation (NNSE). The stopping criteria was also set to 1% improvement. A sensitivity analysis was finally carried out to show the importance and effect of each input variable in the performance of the developed model. Based on this, the model was run with one input removed and the NNSE and RMSE were both observed. Furthermore, an uncertainty analysis was carried out for one step ahead to show how relative accuracy is changed when running the model with the portion of the data used for model development. The model was run by 5% reduction each time and the new performance was reported. This is also shown in Figure 6 under the result and discussion section.

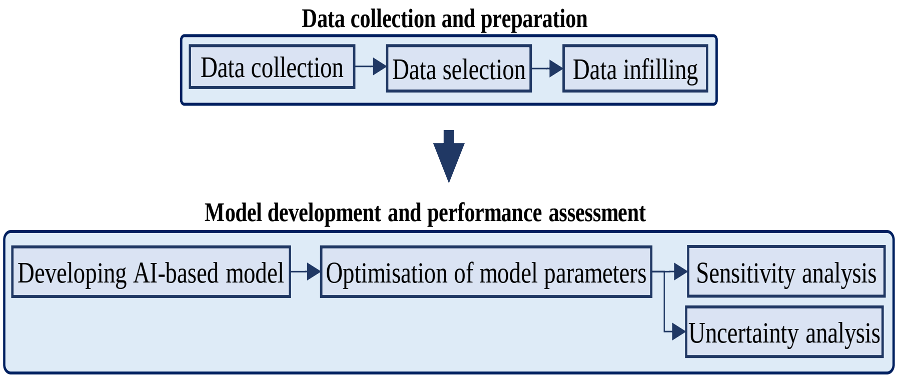


Figure 1. Schematic framework of the methodology for biogas prediction in anaerobic digestion systems

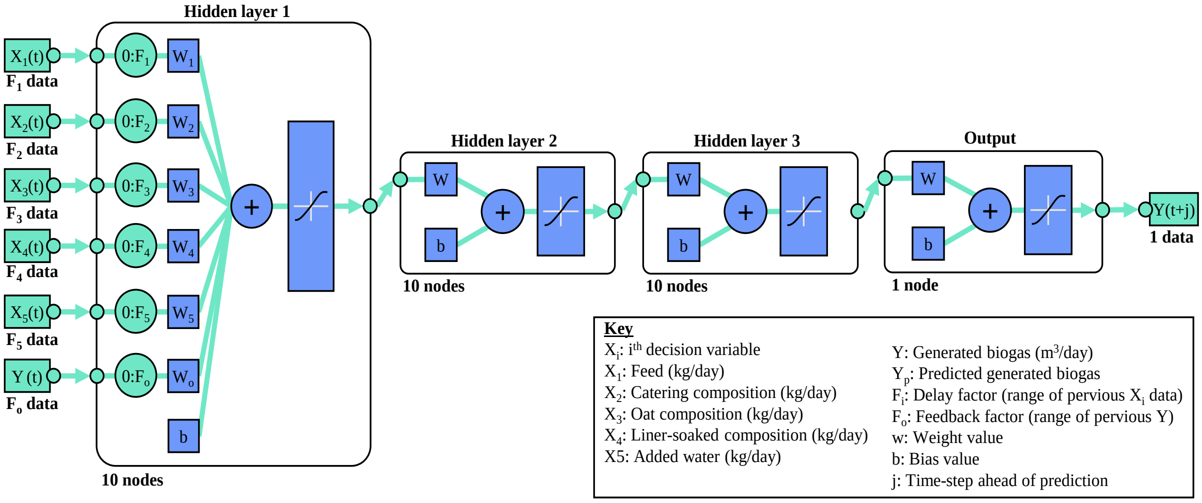


Figure 2. Structure of the proposed NARX model

3. Results and Discussions

The results presented in this section were obtained following the outlined description in the methodology. This is in line with Figures 1 and 2 which shows the schematic framework for the prediction of biogas from an anaerobic digestion system and the structure of the proposed NARX Model.

Figure 3 shows the performance of the tests carried out for each of the different data mining techniques using the Root Mean Square Error (RMSE) metric to assess each of the technique. This was to determine the most appropriate method used to infill the missing biogas data. As can be seen in the results, it was observed that the Kriging Technique had the least range of fluctuations amongst other data mining techniques as it had an average RMSE value of 1.23 cubic meters/day while others had average RMSE values ranging from 1.25-2.25 cubic meters/day. Based on this observation, the Kriging Technique was then selected and used to obtain the missing biogas values which had complete feed values.

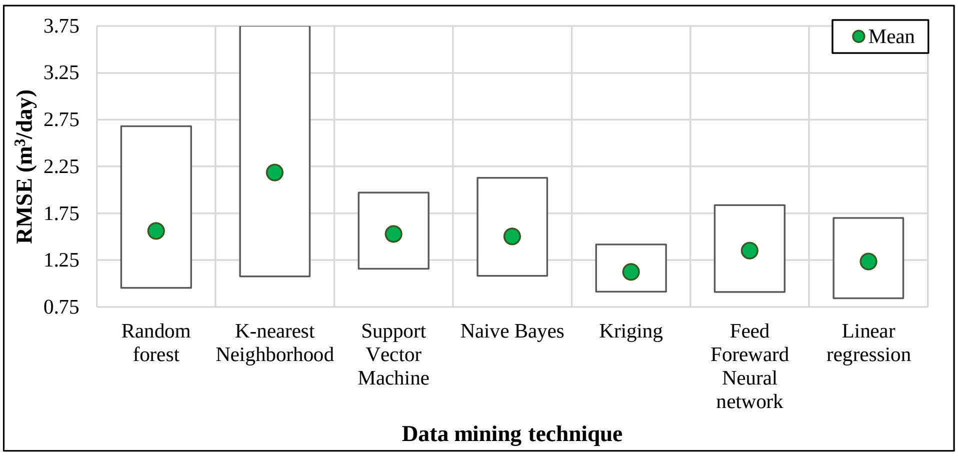


Figure 3. Performance of data mining techniques applied for infilling missing data

As NARX model needs specifying lag times for each input (decision variables), So the optimisation method is developed to find best lag times for each decision variable. The optimisation model was developed, and the results were converged and obtained after 8 iterations as the results of the best lag times can be seen in Table 1. Figure 4 shows the results of the performance of the optimisation of the NARX Model which were assessed using the Root

Mean Square Error (RMSE) and Neural Network-Based State Estimation (NNSE). From the results in Figure 4, the RMSE was observed to decrease for each trial step ahead while the NNSE increased concurrently. Table 1 shows the results of the best decision variables obtained for each trial of the optimised model where the catering composition, oat composition and liner-soaked composition had a lag time of 1 which was equivalent to the generated biogas. This indicates that the catering composition, oat composition and liner-soaked composition are heavily dependent on the generated biogas compared to the other input variables. Lag times also show that biogas is correlated by 3 days before of feeding and 5 days before of added water but added waste (composition) correlated by only one day before.

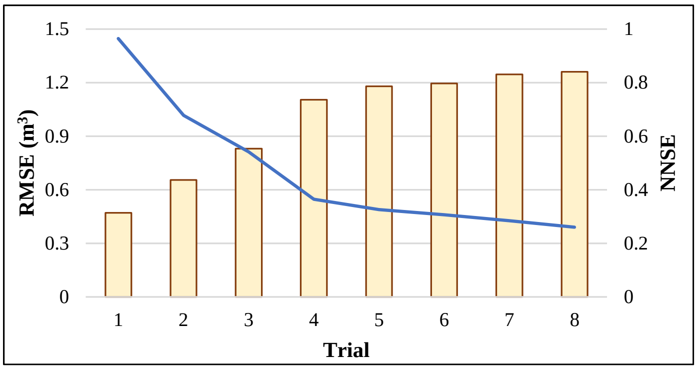
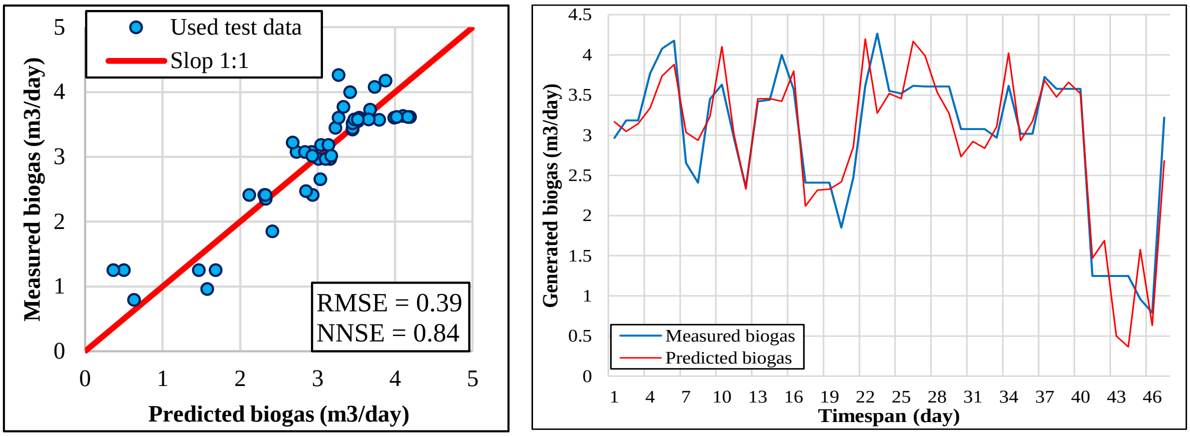


Figure 4: Metrics results of model improvement using optimisation model

Decision variable	**Lag time in each trial**							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Feed (F_1)	5	6	4	7	10	3	3	**3**
Catering composition (F_2)	3	1	1	1	1	1	1	**1**
Oat composition (F_3)	3	1	1	1	1	1	1	**1**
Liner-soaked composition (F_4)	3	1	1	1	1	1	1	**1**
Added water (F_5)	2	2	3	3	3	4	4	**5**
Generated Biogas (F_o)	3	1	2	7	7	5	11	**1**

Table 1. Best decision variables for each trial of the optimised model

Figure 5 shows the performance assessment of the developed model where both the scatter plot of predicted biogas vs corresponding measurements for 1 day ahead and the comparison of observations with the predictions were observed. From Figure 5a, the NNSE value obtained was 84% (0.84). This indicates that the time efficiency of the model developed is relatively high. It also indicates that the developed model was robust compared to the conventional weighted least squares state estimated method. The RMSE value obtained was 0.39 which was observed to be quite low considering the number of datasets used to develop the model. This indicates that the difference between the measured biogas values and the predicted biogas values were quite insignificant as shown in Figure 5b. This thereby confirms the effectiveness of the developed model.

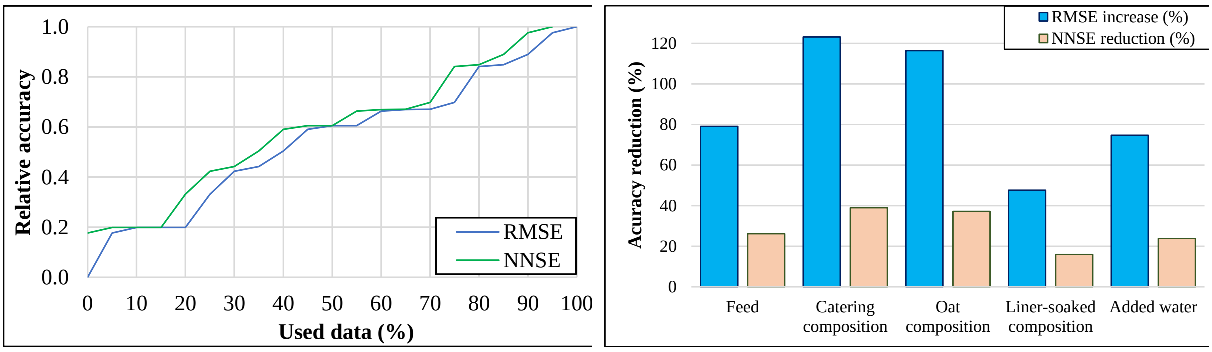


(a) **(b)**

Figure 5: Performance assessment of developed model: (a) Scatter plot of predicted biogas vs corresponding measurements for 1 day ahead, (b) Comparison of observations with the predictions

Figure 6 presents a practical illustration of the further analysis carried out on the developed model where the results of both the uncertainty and sensitivity analysis were presented. The uncertainty analysis presented in Figure 6a showed a decrease in both the RMSE and NNSE following a decrease in the percentage of used datasets. This implies that prediction accuracy of the NARX Model decreases with a corresponding decrease in the dataset. It also indicates that the prediction accuracy of the model is higher with a higher number of datasets.

The sensitivity analysis was conducted by removing one decision variable and running model for only one step ahead. From the sensitivity analysis presented, it was observed that the model was heavily dependent on the oat and catering composition as they both demonstrated the highest levels of accuracy compared to the other input variables. This indicates that the oat and catering composition have the greatest significant impact on the performance of the model compared to the other input variables. However, the feed and water added to the digester also demonstrated to have significant impacts on the model performance compared to the liner-soaked composition which was the least as shown in Figure 6b.



(a) **(b)**

Figure 6: Further analysis: (a) uncertainty analysis on percentage of used dataset for development of the model, (b) sensitivity analysis on role of decision variables in accuracy of developed model

4. Conclusions

From the results obtained and the discussion made on each of the results, it can be deduced that the developed NARX Model gave accurate predictions of the biogas generated from the decentralised AD plant as the RMSE was relatively low and the NNSE relatively high considering the number of datasets used in developing the model. This implies that the NARX Model can give good and reliable predictions with little sets of data. In addition, the performance of the NARX Model is highly dependent on the oats and catering composition compared to the feed, soaked-liner composition and water added to the digester. This implies that the production of biogas from the decentralised AD plant was greatly influenced by the composition of these two feedstocks.

Furthermore, the results obtained from the development of the NARX model for real time monitoring of a decentralised Anaerobic Digestion (AD) System indicates that the NARX model can address the technical limitations of the AD system. Through this means, the efficiency and effectiveness of the AD system in generating biogas will be greatly improved. The resultant effect of this will facilitate the implementation of circular economy and most importantly enable the Anaerobic Digestion technology in achieving the Sustainable Development Goals (SDGs) set out by the UN to achieve by 2030 which has been revealed to be linked to 16 out of the 17 Sustainable Development Goals.

5. Acknowledgement

This work is supported by the Knowledge Exchange (KE) Seed Fund allocated to the fifth author (industrial partner) and the Fellowship allocated to the third author. The authors wish to acknowledge the KE seed fund supported by the University of West London and the Fellowship supported by the Royal Academy of Engineering under the Leverhulme Trust Research Fellowships scheme. The authors also wish to thank the Diego Vega from LEAP Ltd and Dr Davide Poggio from The University of Sheffield for their great support to provide and analyse the data collected from the case study.

6. References

Arun.C., S. 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*, 200-210.

Association, W. B. (2021). *Biogas: Pathways to 2030.* Retrieved from worldbiogasassociation.org

Bahreini, G., Nazari, L., Ho, D., Flannery, C.C., Elbeshbishy, E., Santoro, D., Nakhla, G., 2020. Enzymatic pre-treatment for enhancement of primary sludge fermentation. Bioresource. Technol. 305 (February), 123071. http://dx.doi.org/10.1016/j.biortech.2020.123071

Bui Quang., N. Q.-H.-M. (2020). Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. *Journal of Hydrology 581:124379*.

Felix Mayer, R. B. (2019). Critical review on life cycle assessment of conventional and innovative waste-to-energy technologies. *Science of the Total Environment 672*, 708-721.

Hao-nan Guo, S.-b. W.-j.-t. (2021). Application of machine learning methods for the prediction of organic solid waste: A Review. *Bioresource Technology 319*.

J.Taherzadeh., S. W. (2020). Resource Recovery and Circular Economy from Organic Solid Waste using Aerobic and Anaerobic Digestion Technologies. *Bioresource Technology*.

Laurent.A., B. J. (2014). A Review of LCA studies of solid waste management systems-part 1: lessons learned and perspectives. *Waste Management 34*, 573-588.

Luis Alberto Bertolucci Paes, B. S. (2019). Organic solid waste management in a circular economy perspective – A systematic review and SWOT analysis. *Journal of Cleaner Production 239*.

Management, C. I. (2021). *Waste Management and Climate Change.* London.

Park, C. L. (2005). Upgrading of anaerobic digestion by Incorporating two different hydrolysis processes. *Journal of Bioscience and Bioengineering. 100*, 164-167.

RecyclingInside. (2021). *Converting Waste to Energy through Incineration: A small business solution.* Retrieved from recyclinginside.com

Shahsavar, M.M., Akrami, M., Gheibi, M., Kavianpour, B., Fathollahi-Fard, A.M. and Behzadian, K., (2021). 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,* p.114794

Shyamala, D. a. (2012). Studies on Variations in Physico-Chemical and Biological Characteristics at Different Maturity Stages of Municipal Solid Waste Compost. *International Journal of Environmental Sciences, 2*, 1984-1997.

Triassi, M. A. (2015). Environmental Pollution from illegal waste disposal and health effects: A review on the "triangle of death". *International Journal of Environmental Research Public Health 12*, 1216-1236.

Wright, S. (2020, October 8). *Part 1: What is an Incinerator?* Retrieved from www.inciner8.com