

1 **Development of an Artificial Intelligence-Based Framework for** 2 **Biogas Generation from a Micro Anaerobic Digestion Plant**

3 Ikechukwu Offie¹, Farzad Piadeh¹, Kourosh Behzadian^{1*}, Luiza C. Campos², Rokiah Yaman³

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

5 ² Civil, Environmental and Geomatic Engineering, University College London, Gower St, London WC1E6BT,
6 UK

7 ³ Leap AD Micro Company, London, UK

8 * Corresponding author. Tel.: +44 (0) 20 8231 2466 E-mail address: kourosh.behzadian@uwl.ac.uk

9 **Abstract**

10 Despite the advantages of the Anaerobic Digestion (AD) technology for organic waste
11 management, low system performance in biogas production negatively affects the wide spread
12 of this technology. This paper develops a new artificial intelligence-based framework to predict
13 and optimise the biogas generated from a micro-AD plant. The framework comprises some
14 main steps including data collection and imputation, recurrent neural network/ Non-Linear
15 Autoregressive Exogenous (NARX) model, shuffled frog leaping algorithm (SFLA)
16 optimisation model and sensitivity analysis. The suggested framework was demonstrated by
17 its application on a real micro-AD plant in London. The NARX model was developed for
18 predicting yielded biogas based on the feeding data over preceding days in which their lag
19 times were fine-tuned using the SFLA. The optimal daily feeding pattern to obtain maximum
20 biogas generation was determined using the SFLA. The results show that the developed
21 framework can improve the productivity of biogas in optimal operation strategy by 43%
22 compared to business as usual and the average biogas produced can raise from 3.26 to 4.34
23 m³/day. The optimal feeding pattern during a four-day cycle is to feed over the last two days
24 and thereby reducing the operational costs related to the labour for feeding the plant in the first
25 two days. The results of the sensitivity analysis show the optimised biogas generation is

26 strongly influenced by the content of oats and catering waste as well as the optimal allocated
27 day for adding feed to the main digester compared to other feed variables e.g., added water and
28 soaked liner.

29 **Keywords:** Anaerobic digestion; Artificial intelligence framework; Biogas generation;
30 Optimised operation strategy; Organic waste; Recurrent neural network.

31 **1. Introduction**

32 Over the years, the world has been subjected to unprecedented population growth, economic
33 development, and rapid urbanisation. These series of development have given rise to a constant
34 increase in organic waste generation globally. This constant increase in the generation of
35 organic wastes has become a major source of concern globally following its negative impacts
36 (Arun and Sivashanmugam, 2017). Organic wastes account mainly for approximately 105
37 billion tonnes of the total municipal solid waste generated on an annual basis globally (WBA,
38 2021). Lack of proper and efficient waste management strategies can lead to a series of
39 environmental problems such as emerging pollution, ecosystem destruction, harm to human
40 health and depletion of natural resources (Kumar *et al.*, 2021). The poor management of
41 organic wastes also has the potential to contribute to climate change through the emission of
42 greenhouse gases into the atmosphere (CIWEM, 2021). The effect of this has compelled
43 nations and governments to invest more financial and material resources for the remediation of
44 organic wastes in recent years (Wainaina *et al.*, 2020).

45 Presently, efforts are being made to revolutionise the waste management industry towards
46 achieving sustainability and profitability (Abdallah *et al.*, 2020). This has led to the application
47 of advanced recycling technologies such as anaerobic digestion (AD), composting and
48 incineration amongst others in treating and managing wastes that having been identified to be
49 better alternatives to landfill systems (Wainaina *et al.*, 2020). AD technology has been regarded
50 as an established biological processing technique suitable for stabilising a plethora of organic
51 solid wastes that also results in resource recovery of energy (i.e., methane biogas) and useful
52 nutrients (i.e., organic fertilisers) (Wainaina *et al.*, 2020). More specifically, the AD technology
53 can deliver both de-fossilisation and decarbonisation, i.e., avoided GHG (greenhouse gases)
54 emissions through converting organic wastes to (1) renewable energy thereby reducing the
55 need for fossil fuel utilisation and (2) organic fertilisers reducing the need for chemical

56 fertilisers (WBA., 2021). In addition, compared to other technologies such as incineration that
57 may result in air pollution and GHG emissions, the ability of the AD technology for converting
58 waste to useful energy and organic nutrients without causing any form of environmental
59 pollution, i.e., avoided embodied energy/carbon and hence avoided GHG emissions, makes it
60 a preferable option (Liang *et al.*, 2022). The multi-faceted nature of the AD technology has
61 rendered it as a highly ranked technique in the waste management industry and an excellent
62 tool for the realisation of circular economy (WBA., 2021).

63 Evidently, the performance of the AD technology is mainly evaluated based on the biogas
64 generation as the most valuable output which is the result of processes in four stages including
65 hydrolysis, acidogenesis, acetogenesis, and methanogenesis (Shahsavari *et al.*, 2021). Despite
66 the plethora of advantages in the AD technology, its performance especially for biogas
67 generation is heavily dependent upon the balanced mix of the waste and microbial groups and
68 hence is highly sensitive to organic compounds and may result in process instability and failure
69 (Cruz *et al.*, 2022). In addition, long residence time and low removal efficiency of organic
70 compounds are other limitations that hinder the wide application and adoption of this
71 technology to full potential (Xu *et al.*, 2021). All this can directly affect the efficiency of the
72 biogas production. Therefore, modelling AD processes are of paramount importance and useful
73 tool to first estimate and then optimise the AD performance (i.e., projection of biogas
74 production and organic fertilisers). Although several conventional mathematical models (e.g.,
75 theoretical, analytical and statistical) are available, their application is limited due mainly to
76 the complexity of their development, data demanding and challenges with model calibration
77 (Cruz *et al.*, 2022). Hence, these models are widely used as useful tools for the AD planning
78 and design such as AQUASIM, GRAINIT BIOGAS, ANESSA and ADM1 (Carlini *et al.*,
79 2020). However, the reliability of these models within the operation phase of AD plants is more
80 challenging as the operation conditions of AD processes can be highly variable and rapid

81 changes in control parameters are inevitable especially depending on waste composition
82 (Cheela *et al.*, 2021). As a result, due to changes in various microbial species and the complex
83 metabolic pathways, the above mathematical models are unable to properly estimate the model
84 performance. However, data driven models such as Artificial Intelligence (AI) can be
85 introduced as a good surrogate for process-based modelling that are dependent of complex
86 physico-chemical processes. In other words, the AI-based models are developed based on
87 historic data of the system variables and can be used for real-time operation of AD plants by
88 using online data (Piadeh *et al.*, 2022).

89 Several research works have studied the application of AI methods to the AD processes for
90 modelling the relevant non-linear and complex relationships by focusing on optimising particle
91 size of organic matters, organic loading rate, ratio of carbon to nitrogen (C/N), pH and
92 temperature, and residence time (Zhang *et al.*, 2019). These research studies mainly followed
93 three approaches: (1) using classification machine learning (ML) methods such as support
94 vector machine, random forest (RF), K-nearest neighbourhood (KNN) to predict the corrected
95 operation, (2) optimising parameters by particle swarm and genetic algorithm, and (3)
96 employing various artificial neural networks (ANN) to predict control parameters (Cruz *et al.*
97 2022). To increase the rate of biogas yield, AI-based methods have been widely used in
98 agricultural and industrial application (Kunatsa and Xia, 2022). However, to the best of our
99 knowledge, few research works have presented an AI-based framework for developing
100 operation strategies to improve the AD performance in producing biogas from the food waste
101 generated in an urban area. More specifically, the KNN method employed by Wang *et al.*
102 (2020) and RF used by Long *et al.* (2021) separately classify and find the regression between
103 different operational control measurements and biogas generation. Tufaner and Demirci (2020)
104 used simple ANN to predict biogas generation in a laboratory scale AD by using pH, alkalinity,
105 organic load rate, chemical oxygen demand (COD) and total solid (TS) Park *et al.* (2021)

106 similarly used pH, alkalinity, COD removal and volatile solids as input variables for ANN to
107 predict biogas yield. More recently, Pei *et al.* (2022) used data mining and ANN models to
108 estimate biogas generation based on TS, C/N, pH and acid concentration. These efforts aimed
109 at estimating the AD outputs especially biogas production based on the system variables
110 especially pH, alkalinity, and effluent pollution. Although the development of smart and
111 decision-making frameworks in waste management have recently attracted more attention by
112 researchers (Shahsavari *et al.*, 2021, Shahsavari *et al.*, 2022), none of the previous works either
113 developed a framework for the AD operation based on the ANN models or carried out proper
114 investigations on the effect of different waste compositions and the water added to the AD on
115 biogas yield. Furthermore, those previously developed models mainly used simple ML or ANN
116 whereas the performance of the AD procedure may fit in better with simulation of time-series
117 models that rely on earlier timesteps. This is particularly important because AD systems are
118 operated continuously and are highly dependent on sequential and continuous input waste load
119 (Yang *et al.*, 2022; Chozhavendhan *et al.*, 2023). This type of modelling can be envisaged
120 through the application of a recurrent neural network (RNN) model for monitoring the
121 performance of the AD processes (Offie *et al.*, 2022). Hence, this study aims to develop a new
122 smart framework for optimal operation performance of micro-AD plants located in a residential
123 area based on Recurrent Neural Network (RNN) and optimisation techniques. It is also aimed
124 at determining the maximum volume of biogas that can be generated from the micro-AD plant.
125 This framework is demonstrated by its application to historic data obtained from a real case of
126 a micro AD plant in London, UK.

127 This paper is organised as follows: in section 2, the features of the micro-AD plant used as the
128 pilot study as well as the description of the micro-AD site location will be clearly stated. The
129 nature of data collected from the micro-AD plant and different techniques adopted for data
130 imputation will be then presented. In addition, the type of artificial neural networks (ANN)-

131 based model developed for monitoring and improving the efficiency of the micro-AD plant
132 will also be presented and described in detail alongside the sensitivity and uncertainty analysis
133 carried out to assess the performance of the developed ANN model. The results obtained from
134 infilling the missing data and the ANN model development and testing will be presented and
135 discussed in detail in section 3 followed by finally summarising key findings and remarking
136 notes in section 4.

137 **2. Methodology**

138 This study presents a new AI-based framework for the simulation and optimisation of micro-
139 AD plants based on data-driven models. Figure 1 shows the methodology in this study
140 comprising three main steps as data collection/preparation, model development and
141 performance assessment. These steps are commonly used for developing most data-driven
142 environmental models (Piadeh *et al.*, 2022). The AI-based framework is mainly used as the
143 core tool for estimating and optimising biogas generation based on the feed data collected over
144 preceding days. All steps of the framework were carried out using MATLAB 2021b software
145 which provides functions for estimation and optimisation of the system performance. These
146 steps follow a series of procedures after collecting data from the micro-AD plant, which are
147 described below with more details.

148 **2.1 Data collection and preparation**

149 This stage entails data collection and imputation for infilling missing data using some data-
150 mining-based techniques, selection of relevant data for model development. The data in this
151 study was collected from a micro-AD plant located in Camley Street Natural Park Central
152 London, United Kingdom (UK) with the schematic diagram shown in Figure 2 (Walker *et al.*,
153 2017). The micro-AD plant in this site had a pre-feed tank consisting of a chopper mill, mixing
154 pre-feed tank on load cells and a feed pump. It also had a main anaerobic digester containing

155 an automated mechanical mixer and heater by an internal water heat exchanger. Other main
156 components of the micro-AD plant as shown in the figure include the hydrogen sulphide
157 scrubber filled with activated carbon pellets, floating gasometer for biogas storage, digestate
158 sedimentation tank, digestate liquor storage tank. The micro-AD plant was monitored for a
159 period of 310 days during which the operational parameters, biological stability, and energy
160 requirements of the micro-AD plant were evaluated.

161 The data collected from the micro-AD plant include temperature, pH, volatile solids, total
162 solids, feed into the main digester, feed composition into the pre-feed tank. The feed
163 composition comprises apple, catering and coffee, coffee, digestate, green waste, oats, soaked
164 peanuts and muesli, tea, tea leaves, tea bags, oil, soaked muesli, soaked liners, and catering.
165 Other data collected are the water added to the pre-feed tank and the volume of biogas
166 generation. The feed into either the pre-feed tank or the main digester was usually done every
167 few days when both feed amounts and biogas volume in the storage were recorded. Hence, out
168 of the monitoring period of 310 days of the micro-AD plant, there were days when no feed was
169 added to either the pre-feed tank or the main digester and no recording of biogas generation
170 while daily continuous data for both feed and biogas are necessary for developing a time-series
171 ANN model that considers lag days. In addition, there were some days with missing output
172 data (i.e., there was feed but there was no reading for biogas generation). This effect can hinder
173 the model accuracy of the micro-AD plant especially for the prediction of the biogas volume
174 generated. Hence, some data-mining techniques were first analysed in this study for estimating
175 the missing data to determine the most suitable one for infilling the missing data. Note that
176 missing data in this study refer to the absence of biogas readings in two types: (1) data samples
177 with feed values available (input) but no reading for biogas generation (output); and (2) data
178 samples with feed value equal to zero but no reading for biogas generation. Therefore, the
179 entire dataset was first divided into two groups of data with feeding inclusive and data without

180 feeding. Some data mining techniques were then tested to identify the relationship between the
181 feed data and the generated biogas for data groups with feeding data. Out of those techniques,
182 the best one was selected for infilling the missing data of the first type (i.e., data with feed
183 values but no biogas values). The second type of missing data (i.e., data where feeding is zero
184 and biogas is unavailable) were infilled based on the linear regression of the remaining total
185 biogas data read. The data mining techniques explored here include Random Forest (RF), K-
186 Nearest Neighbour (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), Kriging, Feed
187 Forward Neural Network (FFNN) and Linear Regression (LR).

188 Sequel to this, a sensitivity analysis was carried out for each of the operational feed variables
189 to determine their correlation and impact on the volume of biogas generation. Based on the
190 cross-correlation analysis of all input variables (demonstrated as Figure A1 and Table A1 in
191 the Appendix), the daily feed into the digester, the water added to the digester showed
192 significant correlation and corresponding impact on the biogas volume generated. In addition,
193 out of various waste compositions, only oats, soaked liners, and catering were selected whereas
194 other waste compositions were negligible as they had no significant correlation and hence no
195 meaningful impact on the volume of the biogas generated. In addition, the volatile solids, total
196 solids, pH, and temperature were measured but observed to be relatively constant during the
197 operation and hence these parameters were also excluded from the analysis for estimating
198 biogas generation.

199 **2.2 Model development**

200 To estimate biogas generation, a type of time-series RNN model known as Non-Linear
201 Autoregressive Neural Network (NARX) was developed here with three hidden 10-neuron
202 layers with the architecture shown in Figure 1. This model was developed based on the selected
203 input variables of the micro-AD plant including the actual/estimated daily feed added to the

204 main digester (X_1), the feed composition comprised of catering (X_2), oats (X_3), and soaked
205 liners (X_4) added to the pre-feed tank (i.e. the top three highly correlated variables with biogas
206 generation), the water added to the pre-feed tank (X_5), and the volume of biogas generation
207 (Y). The model settings are as follows: Levenberg-Marquardt method used for training process;
208 mean square error as the indicator to evaluate the model performance, and 6 epochs (iterations)
209 adjusted for training failure. The database used for model development is divided into three
210 parts as 70% for training, 15% for validation and 15% for test as a common practice (Eghbali
211 *et al.* 2017). The trained model was then used to predict the biogas generation (Y_p) for the
212 micro-AD plant in the case study.

213 As the NARX model needs lag time specification in day (known as delay factor F_i), i.e., range
214 of input variables for previous timesteps to use for estimation of biogas generation at one
215 timestep ahead (Y_{t+1}) based on input data (decision variables), an optimisation method is used
216 to find the optimal lag time for each decision variable, as a model tuning, to obtain the most
217 accurate output data i.e., biogas generation. To this effect, the optimisation model was
218 developed using shuffled frog leaping algorithm (SFLA). This is a memetic and nature-based
219 algorithm with the ability to search in both local and global search space where each lag time
220 represents one frog (Bui *et al.*, 2020). Here, each frog, i.e., decision variable, represents a lag
221 time to find the minimum root mean square error (RMSE) and the highest Normalized Nash-
222 Sutcliffe Efficiency (NNSE) in this optimisation approach. 4 trials for exploration and 4 trials
223 for exploitation were set for each iteration of optimisation, and stopping criteria being set to an
224 improvement of less than 1%. Each of the six decision variables (i.e., F_0 - F_5 in Figure 1) is an
225 integer value ranging between 0 and 10 due to the results of cross-correlation analysis on inputs,
226 provided in Figure A1 in the appendix. Thus, this approach can be used to determine the delay
227 factor (range of previous X_i data) for each input data/decision variable.

228 This algorithm was then used again to specify the required weights for the daily feeds added to
229 the main digester, daily feed compositions and the water added to the pre-feed tank to maximise
230 the output (i.e., maximum volume of biogas generation from the micro-AD plant) for each of
231 the days in a cyclic period of feeds. Note that the cyclic period is based on the (lag time) delay
232 factor specified in the first optimisation model. While stopping criteria and trials are set
233 similarly, each NARX input for each day are selected as decision variables. To simulate the
234 real operation and put a cap for the feeds/water added to the plant, constraints are defined based
235 on the historic operation of the plant as follows: (1) maximum feed equals to 80 kg every 4
236 days. Note that 4 days is based on the cyclic period of 5 days (as four days input data and biogas
237 generation in day 5) specified as a result of the optimisation model for the largest lag time (see
238 the result section); (2) total weight of the feed and all pre-feeding compositions should be
239 equalled during the optimisation; (3) added water is limited to 30% of the total feed weight, (4)
240 all decision variables need to be either zero or positive values.

241 **2.3 Performance assessment**

242 Two metrics i.e., RMSE and NNSE are used in this study to evaluate the performance of the
243 developed NARX model. The developed model is also evaluated using both the sensitivity
244 analysis and uncertainty analysis. The sensitivity analysis is carried out to show the
245 significance of each input variable for yielding the predicted output (biogas) by removing one
246 input parameter and running the model afterwards with NNSE and RMSE both for
247 observations. The uncertainty analysis on the other hand was done to show how the relative
248 accuracy is changed when running the model with dataset reduction. When carrying out the
249 sensitivity analysis, the optimal waste composition is taken into consideration where the impact
250 of each waste composition is analysed and evaluated to determine the significant impact of
251 each waste composition on the generation of biogas. Further analyses are also carried out on
252 the feed data to further evaluate its importance in the performance of the developed model.

253 **3. Results and discussion**

254 **3.1. Data imputation results**

255 Figure 3 shows the performance of the data mining techniques applied for infilling the missing
256 data. The performance of these techniques is shown for RMSE of the test data based on the
257 cross-validation method in which all data samples participate in the evaluation of the test set
258 (Eghbali *et al.*, 2017). The 6-fold cross-validation method was used in this study for the
259 performance assessment of infilling the missing data. Based on the results presented, it is
260 evident that the Kriging technique had the least range of fluctuations amongst other data mining
261 techniques with an average RMSE value of 1.23 m³/day compared to 1.25-2.25 m³/day for the
262 other techniques. As a result, the Kriging technique was selected and used to obtain the missing
263 biogas values with complete feed values thereby giving rise to more accurate predictions of the
264 biogas produced from the AD plant. This further confirms the effectiveness of the Kriging
265 technique in infilling missing data where other previous studies have not highlighted dealing
266 with missing data and usually have used simple techniques, especially linear regression (Pei *et*
267 *al.*, 2022). This is specifically important because similar to many industrial and real practices,
268 there was no daily measurement for the generated biogas and only 123 non-sequential data out
269 of the 310 operation days were recorded. Hence, this infilling technique could provide
270 acceptable results used to develop the model for accurate estimation and optimisation of the
271 generated biogas volume.

272 **3.2. Fine-tuning of lag times of input variables**

273 Figure 4 shows the optimal number of lag times for each input variable is obtained after 8 trials.
274 This figure also shows the obtained lag times in each iteration and their corresponding model
275 performance metrics (i.e., RMSE and NNSE). As can be seen in Figure 4, the RMSE and NNSE
276 were observed to decrease and increase gradually over the trial number, respectively. The
277 results show that the optimal daily lag time data is 5 days for the added water (F₅), followed by

278 3 days for the feed added to the main digester (F_1), and then 1 day for other variables (i.e.,
279 catering, oat, soaked-liner, and biogas generation). This shows that adding water to the pre-
280 feed tank can influence the content of the generated biogas in the main digester from the past
281 5 days whereas waste compositions can immediately impact on biogas generation from only
282 one previous day. This also shows the yielded biogas can be dependent on the daily distribution
283 of feeds and water. Furthermore, one day lag time in the waste composition indicates that the
284 process of biogas generation is highly influenced by specific rate of waste compositions added
285 to pre-digester, even if this rate is different from that of material added the digester.

286 Furthermore, the cross-correlation analysis, provided in Figure A1 in the appendix, shows the
287 highest correlation coefficient between biogas generation and the feed adding to the main
288 digester occurs in previous 5 days (used as initial for F_1 in Figure 4) whereas optimised lag
289 time for feed is reduced to 3 days (trial 7 in Figure 4). Similarly, daily lag times for catering,
290 oat and liners are reduced from 3 days in initial trial based on the cross-correlation analysis to
291 only 1 day. However, the high correlation of 5th to 3rd days ago for the added water were initially
292 ignored (number 2 in the initial row for F_5 in Figure 4 vs number 5 in the last row). This can
293 be due to the impact of the combination of input variables on optimal lag times that is shown
294 in the significant improvement of metrics, i.e., RMSE (decrease from 1.4 to 0.4) and NNSE
295 (increase from 0.6 to around 0.9). Although most of the previously developed NARX models
296 recommended using cross-correlation results for developing NARX model (Abdel daiem *et al.*,
297 2022), the difference between initial lag times and final lag times obtained from the SFLA
298 method shows the added value of using optimisation models to fine tune these time-series
299 models.

300 **3.3. Performance assessment of biogas predictions**

301 Figure 5a compares the biogas measurements with the corresponding estimations over the test
302 period. From this figure, the disparity between the predicted biogas and the generated biogas
303 was observed to be quite insignificant. Figure 5b also shows the performance assessment of
304 the developed model where the scatter plot of predicted biogas versus corresponding
305 measurements for one day lead time (i.e., one day ahead) for the three types of the feed. The
306 RMSE values are observed to be 0.33 m³/day for heavy weight feed (greater than 20 kg), 0.46
307 m³/day for medium feed (within 10-20 kg) and 0.39 m³/day for light weight feed (less than
308 10kg). While it can be deduced that the feed with weight greater than 20 kg had the least range
309 of errors compared to other two feeds, the coefficient of variance (CV), indicates that the biogas
310 estimation model is highly sensitive to the feed with lower weights compared to the feed with
311 higher weights. However, the three RMSE values obtained which are relatively low indicating
312 that the efficiency of the model developed is relatively high and hence reliable to be used as a
313 surrogate model for estimation of biogas generation in the micro-AD plant. Furthermore, the
314 coefficient of variance (CV) of 13% in this study (Figure 5b) can be compared to previously
315 reported studies which are in a range between 31% by Wang *et al.*, (2020) and 23% by Long
316 *et al.* (2021). This confirms the effectiveness of the developed model in predicting the biogas
317 generated from the micro-AD plant. It also indicates that the developed NARX-ANN model is
318 robust enough to be used for relatively high fluctuated biogas generation. Although the overall
319 NNSE can be quite acceptable in this model but high variability of the measured biogas
320 (3.26±1.21 reported in Table A1 in the appendix) may be considered as a drawback for the
321 model in tracking biogas, particularly when there is a sudden change (e.g., droop at end days
322 as shown in Figure 5a).

323 Figures 5c and 5d present further sensitivity analysis for the impact of the percentage of the
324 data used and each of the feed type, respectively, on the accuracy of the developed model.

325 Figure 5c shows the prediction accuracy of the model development in both metrics (NNSE and
326 RMSE) is changed with a relatively similar and linear trend for the percentage of the data used.
327 As the limited dataset (only 310 days) was used for all steps of training, validation and testing,
328 the developed model would still be highly dependent on the volume of dataset. This would also
329 confirm a relatively low coefficient of cross-correlation analysis between the input variables
330 and the generated biogas, as illustrated in Figure A1 in the appendix. The sensitivity analysis
331 presented in Figure 5d is also conducted by removing one decision variable and running model
332 for one step ahead. From the sensitivity analysis presented, the catering and oats compositions
333 demonstrate the highest impact on the prediction accuracy compared to other input variables.
334 This also indicates that both the oat and catering compositions have the most significant impact
335 on the biogas generation in the micro-AD plant compared to the other input variables. On the
336 other hand, the liner-soaked composition has the least impact on the prediction accuracy and
337 hence minimum impact on the biogas generation in the waste composition.

338 **3.4.Optimal feeding pattern and maximum potential biogas generation**

339 The SFLA optimisation method is used to specify the optimal feeding pattern to obtain
340 maximum biogas generation. Based on the optimal number of lag times (days) for each variable
341 obtained in Figure 4, the optimisation method is arranged for 18 decision variables as follows:
342 four variables for feed added to the main digester at days t-3, t-2, t-1 and t; six variables for the
343 water added to pre-feed tank at days t-5, t-4, t-3, t-2, t-1 and t; eight variables for each of the
344 three waste composition types (i.e., catering, oat, and liner) and biogas generation at days t-1
345 and t. The objective value is to maximise the biogas generation at day t+1. The SFLA is run
346 and the results of the optimal feeding pattern within the these setting days for each variable
347 along with optimal biogas generation are shown in Table 1. Note that as the values for the
348 added water are zero for all days except day t-1, the optimisation model was run for smaller
349 number of days for the added water and the results showed that maximum biogas is generated

350 when a four-day cycle is considered for feeding pattern. Hence, the optimal decision variables
351 in Table 1 are shown for 4 days between t-3 and t. The estimation of maximum biogas
352 generation at day t+1 is 4.52 m³/day based on the optimal decision variables at the preceding
353 days between days t and t-3. This is the maximum possible biogas generation from this micro-
354 AD plant based on the optimal values of decision variables. The analysis of the optimal decision
355 variables shows that the entire feed for a four-day cycle only needs to be added on the last day
356 (i.e., 80kg on day t). Out of this 80kg feed, catering added to the pre-feed tank needs to be 60kg
357 with a distribution of 55kg at day t-1 and 5kg at day t; the remaining waste added to the pre-
358 feed tank is 20kg oat added only at day t-1; there is no need for adding any liner; and finally,
359 15kg water added to the pre-feed tank is needed only at day t-1. Accordingly, the biogas
360 generation of the following day (i.e., day t+2) is estimated 4.23 m³/day based on the input
361 (decision) variables of the preceding four days (i.e., between days t-2 and t+1). Table 1 shows
362 the summary of the estimated biogas generation but more details of biogas generation for these
363 four days (i.e., between days t+1 and t+4) are given in Table A2 in the appendix. This decrease
364 in biogas generation is because the daily distribution of input especially feed, catering and
365 water is different from the optimal values obtained above. Similarly, estimation of biogas
366 generation in the following days (i.e., day t+3 and t+4) was observed to decrease further. On
367 the other hand, if the same amounts of feed, catering and water are added every 4 days, the
368 estimated biogas generation is repeated every 4 days. In other words, the estimation of biogas
369 generation after day 4 is observed to be repeated as the same for the input variables with the
370 same feeding pattern. This indicates the volume of biogas generation and feeding pattern can
371 be repeated every four days. As there is no feeding in the first two days, this can also be
372 economically beneficial for the operation of the micro-AD plant which can be mainly operated
373 by local communities with minimum labour (i.e., most of the feeding is arranged for one day
374 every four days) to achieve the maximum efficiency of biogas generation.

375 Figure 6 shows the sensitivity analysis for biogas generation at day t+1 based on percentage of
376 variables over the past three days i.e., days t-2, t-1, and t for all decision variables. Figure 6a
377 shows the impact of the "feed to the main digester" on the biogas generated where different
378 percentages of the feed data for days t and t-1 are shown in horizontal axes and feed data for
379 day t-2 are shown as graphs with an interval of 10%. As can be seen, the maximum biogas
380 generation (4.52 m³/day) can only occur when AD is fed only on the last day (day t). In
381 addition, any redistribution of feeding shows a decrease in the biogas generation which can be
382 translated as relative sensitivity of the model to the daily distribution of feed. For example,
383 when feeding the AD plant at day t-1 instead of day t, the biogas generation dropped from 4.5
384 to around 2.5 m³/day (see blue circle in left-top and right-bottom). The results also show that
385 model is highly sensitive to the amount of feed in last two days (i.e., day t and t-1) and feed
386 ratio for day t-2 is less important (See blue circles are varied more than other lines indicating
387 the model is sensitive to day t and day t-1).

388 Figure 6b shows the impact of the water added on the volume of biogas generation with 20%
389 intervals. Compared to feed distribution in Figure 6a, the daily distribution of added water has
390 a relatively low impact on the biogas generation, that slightly changes between 4.3 to 4.5
391 m³/day. Figure 6c presents the impact of the three different composition variables on the
392 volume of biogas generation. It shows the catering composition added to the pre-feed tank
393 results in the maximum biogas generation compared to other variables. This implies that the
394 catering composition has a higher influence on biogas generation than the other composition
395 variables. Following this, the oat composition also generates a high volume of biogas as shown
396 in Figure 6c. This also indicates that the oat composition has a strong influence on the volume
397 of biogas generation. This is also in line with the sensitivity analysis presented in Figure 6d
398 that shows the model is more sensitive to adding liner rather than daily distribution of

399 composition. For example, while daily distribution of oats and catering has no significant
400 impact on the generated biogas, it is highly sensitive to the amount of added liner in Figure 6e.

401 When applying optimal operation strategy, it is crucial to understand the significance of the
402 distribution of composition variables added to the pre-feed tank over the cyclic period. Hence,
403 the impact of distribution of the optimal amount of the pre-feed composition variables on
404 biogas generation is further analysed in Figure 6d-e for three individual variables i.e., catering,
405 oats, and liner separately. More specifically, Figure 6d considers distribution of optimal value
406 of the catering variable at day t where other variables are fixed here as 20kg for oat, zero for
407 liner and optimum condition of accumulative rate of catering is 60kg (i.e., 55kg for day t-1 and
408 5kg for day t). As can be seen in Figure 6d, the biogas generation slightly decreases for other
409 distribution rates down to 4.515 m³. This indicates that while the AD plant is highly dependent
410 on catering (as shown in the sensitivity figures), its distribution between the day t and t-1 has
411 no significant impact on biogas generation. Although there is a drop in the volume of biogas
412 generation as the share of oat on day t increases compared to the optimum value, the drop in
413 biogas volume has no significant impact on biogas generation. In Figure 6e, the share of liner
414 in day t (%) is observed where each line corresponds to a percentage of liner in total waste. The
415 horizontal axis shows how this percentage is distributed between day t and day t-1. Hence, for
416 20% of the liner, the available data is for 0-5-10-15 and 20%. From Figure 6e, it can be
417 observed that the increasing the liner results in a decrease in the biogas volume as the liner has
418 low impact on biogas generation compared to other waste types.

419 Finally, the strategy for generating optimised biogas generation is compared with best feeding
420 events and entire test period (47 days). As it can be seen, the biogas generation in all three best
421 identified feeding events (as shown in Figure 7a) is relatively similar and a uniform increase
422 with a maximum weekly volume of 26.14 m³. However, the maximum volume of biogas

423 generation for the optimised operation is 29.97 m³ i.e., an improvement rate of 15%. In Figure
424 7b, the average daily biogas generated over the first 40 days is observed to 3.26 m³/day. The
425 generated biogas decreases from days 40-46 as the average volume of biogas generation
426 between days 41-46 is observed to be 1.48 m³ /day. Similarly, the generated biogas for the
427 entire test period for the measured event shows an increase up until day 40 when it experienced
428 a slight decrease in the generated biogas volume. It then increases the next day with a maximum
429 volume of 139.51 m³. On the other hand, the generated biogas in the optimised operation can
430 uniformly increase with a steeper slope and achieve up to a maximum of 199.46 m³ i.e., a
431 significant improvement of 43% for biogas generation compared to the business-as-usual
432 operation. The proposed model has better performance in both short- and long-term operation,
433 i.e., 7 days and 47 days which longer period results in more biogas enhancement. This indicates
434 the potential benefit of developing an optimised strategy for the operation of the micro-AD
435 plant that results in maximum biogas generation and hence the improvement of overall
436 performance and productivity of the micro-AD plant.

437 **4. Conclusions**

438 This paper developed a three-step AI-based framework for estimating and optimising biogas
439 generation from the real-world micro-AD plant. The first step entailed data collection and
440 imputation for infilling the missing data by using several methods, in which the Kriging
441 technique outperformed conventional techniques, particularly KNN, SVM, LR and FFNN.

442 A NARX model was developed in the second step for estimation of biogas generation of the
443 following day based on variables of feed and waste composition added to the main digester and
444 the pre-feed tank for the preceding days which were initially determined based on cross-
445 correlation of time-series of biogas generation and the above variables. The SFLA optimisation
446 model was then used to fine tune the number of lag times of the input layer variables. The result

447 of fine-tuning the lag times showed the yielded biogas is highly sensitive to waste composition
448 (catering, oat and liner-soaked) up to only one previous day, whereas this is up to three previous
449 days for the feed and five previous days for the added water.

450 The SFLA optimisation model was used again to determine the optimal daily feeding pattern
451 that generate maximum biogas in the micro-AD plant. The results show that the optimal feeding
452 pattern can increase the biogas generation up to 15% and 43% for a 7-day and 45-day period,
453 respectively with an average RMSE of 0.39 m³/day. The optimal daily feeding pattern is
454 obtained for a four-day cycle in which water, catering and oat are added to the pre-feed tank
455 within the last two days (mainly on day 3) and the feed is added to the main digester in day 4
456 and hence the no feeding is needed within the first two days. This can also reduce the
457 operational costs related to the labour for feeding the plant.

458 The sensitivity analysis of the developed model shows that biogas generation is strongly
459 influenced by the oats and catering content compared to other feed types. In addition, any
460 change to the optimal daily distribution of the feed added to the main digester is much more
461 sensitive to biogas generation compared to other three feed types (i.e., added water, oat, and
462 catering). Hence, to obtain high biogas generation, it is recommended that the micro-AD plant
463 is fed in one day and allowed to rest for three days compared with gradual feeding every day.
464 Furthermore, adding liner to the plant can significantly reduce the volume of biogas generation.

465 Although the results in this study show significant performance improvement of the micro-AD
466 plant in generating biogas, the developed framework needs to be further tested and verified in
467 other AD plants with longer analysis periods to show the efficacy of the developed approach.
468 In addition, further studies need to be carried out for improving the model ability in tracking
469 sudden change of feed as a common challenge for the AD operation in practice. Further analysis

470 and data modelling can be performed to alleviate other challenges in the AD technology such
471 as long residence time.

472 **Declaration of competing interests**

473 The authors declare that they have no known competing financial interests or personal
474 relationships that could have appeared to influence the work reported in this paper.

475 **Acknowledgements**

476 This work is supported by the Knowledge Exchange (KE) Seed Fund allocated to the fifth
477 author (industrial partner) and the Fellowship allocated to the third author. The authors wish to
478 acknowledge the KE seed fund supported by the University of West London and the Fellowship
479 supported by the Royal Academy of Engineering under the Leverhulme Trust Research
480 Fellowships scheme. The authors also wish to thank the Diego Vega from LEAP Ltd and Dr
481 Davide Poggio from the University of Sheffield for their great support to provide and analyse
482 the data collected from the case study. The authors also wish to thank the editor and the three
483 anonymous reviewers for making constructive comments which substantially improved the
484 quality of the paper.

485 **References**

- 486 Abdallah, M., Abu Talib, M., Feroz, S., Nasir, Q., Abdalla, H., Mahfood B. (2020). Artificial
487 intelligence applications in solid waste management: A systematic research review.
488 *Waste Management*, 109, 231-246.
- 489 Abdel daiem, M., Hatata, A., Said, N. (2022). Modeling and optimization of semi-continuous
490 anaerobic co-digestion of activated sludge and wheat straw using Nonlinear
491 Autoregressive Exogenous neural network and seagull algorithm, *Energy*, 241, 122939.
- 492 Ankun Xu, H. C. (2021). Applying artificial neural networks (ANNs) to solve solid waste-
493 related issues: A critical review. *Waste Management* 124, 385-402.

494 Antoniou, N., Monlau., F., Sambistu., C., Ficara., E., Barakat., A., Zabaniotou., A. (2019).
495 Contribution to Circular Economy options of mixed agricultural wastes management:
496 Coupling anaerobic digestion with gasification for enhanced energy and material
497 recovery. *Journal of Cleaner Production* 209, 505-514.

498 Arun, C., Sivashanmugam, P. (2017). Study on optimization of process parameters for
499 enhancing the multi-hydrolytic enzyme activity in garbage enzyme produced from
500 preconsumer organic waste. *Bioresource Technology* 226, 200-210. Association, W. B.
501 (2021). *Biogas: Pathways to 2030*. Retrieved from worldbiogasassociation.org

502 Carlini, M., Castellucci, S., Mennuni, A., Selli, S. (2020). Simulation of anaerobic digestion
503 processes: Validation of a novel software tool ADM1-based with AQUASIM. *Energy*
504 *Reports*, 6(6), pp. 102-115.

505 Chartered Institute of Water and Environmental Management (CIWEM). (2021). *Policy*
506 *Position Statement on Waste and Resource Management*. Available at www.ciwem.org
507 [retrieved 02-11-2022].

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

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

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

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

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

523 Kunatsa, T., Xia, X. (2022). A review on anaerobic digestion with focus on the role of biomass
524 co-digestion, modelling and optimisation on biogas production and enhancement,
525 *Bioresource Technology*, 344(B), 126311.

526 Long, F., Wang, L., Cai, W., Lesnik, K., Liu, H. (2021). Predicting the performance of
527 anaerobic digestion using machine learning algorithms and genomic data, *Water*
528 *Research*, 199, 117182. doi.org/10.1016/j.watres.2021.117182

529 Offie, I., Piadeh, F., Behzadian, K., Alani, A., Yaman, R. and Campus, L., (2022). Real-time
530 monitoring of decentralised Anaerobic Digestion using Artificial Intelligence-based
531 framework, *International Conference on Resource Sustainability (icRS)*, Virtual
532 conference. <http://repository.uwl.ac.uk/id/eprint/9368/1/iCRS%20Paper.pdf> [retrieved
533 02-11-2022].

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

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

539 Pei, Z., Liu, S., Jing, Z., Zhang, Y., Wang, J., Liu, J., Wang, Y., Guo, W., Li, Y., Feng, L.,
540 Zhou, H., Li, G., Han, Y., Liu, D., Pan, J. (2022). Understanding of the interrelationship
541 between methane production and microorganisms in high-solid anaerobic co-digestion

542 using microbial analysis and machine learning, *Journal of Cleaner Production*. 373,
543 133848.

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

549 Shahsavari, M.M., Akrami, M., Kian, Z., Gheibi, M., Fathollahi-Fard, A.M., Hajiaghaei-
550 Keshteli, M. and Behzadian, K., (2022). Bio-recovery of municipal plastic waste
551 management based on an integrated decision-making framework. *Journal of Industrial
552 and Engineering Chemistry*, 108, pp.215-234.

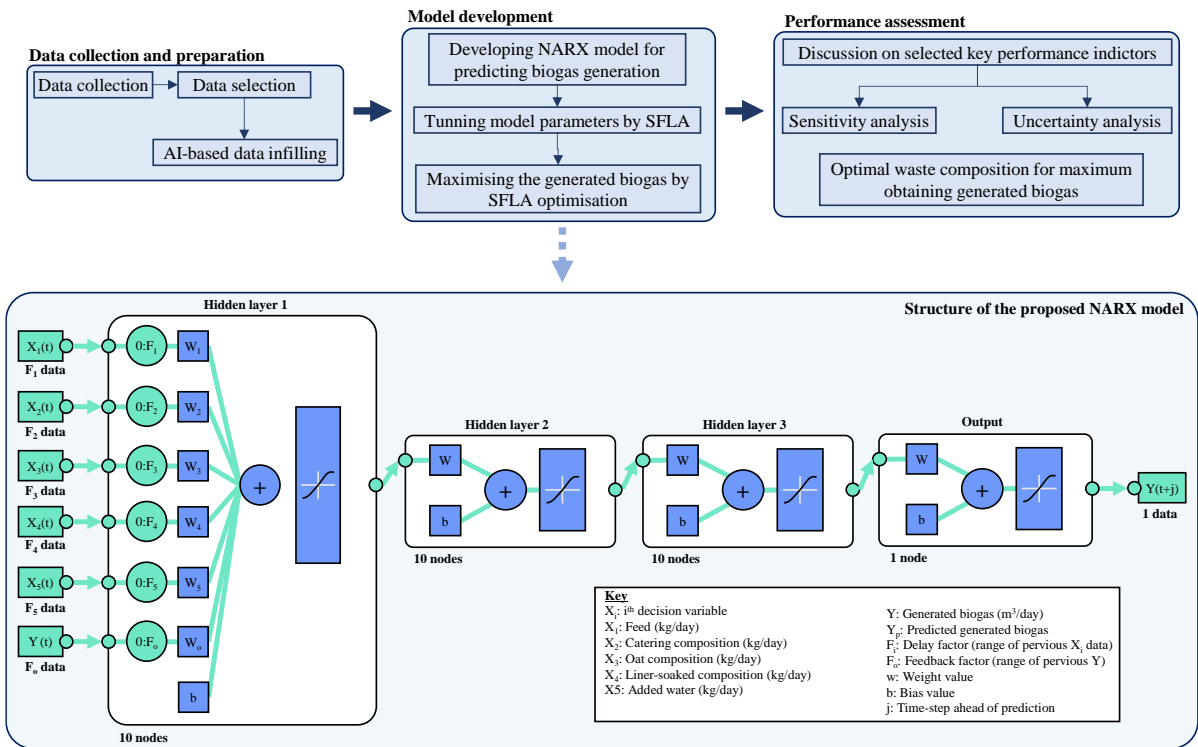
553 Tufaner, F., Demirci, Y. (2020). Prediction of biogas production rate from anaerobic hybrid
554 reactor by artificial neural network and nonlinear regressions models. *Clean
555 Technology Environmental Policy*, 22, pp. 713–724.

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

559 Walker, M., Theaker, H., Yaman, R., Poggio, D., Nimmo, W., Bywater, A., Pourkashanian,
560 M. (2017). Assessment of Micro-Scale Anaerobic Digestion for Management of Urban
561 Organic Waste: A Case Study in London, UK. *Journal of Waste Management Journal
562 of Waste Management* 2017.01.036.

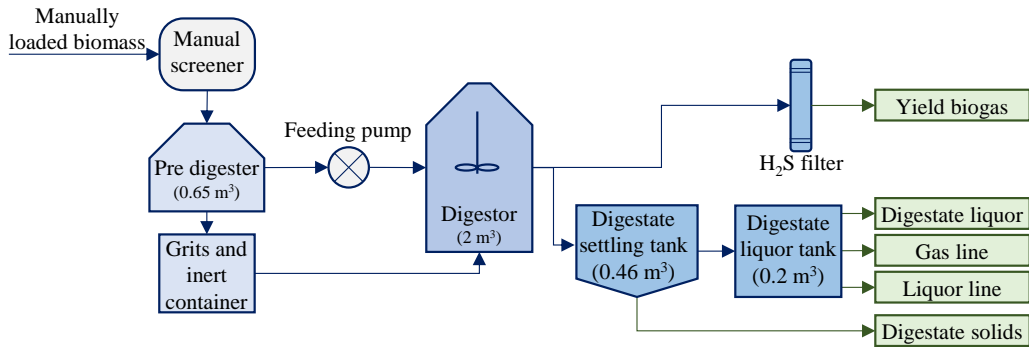
563 Wang, L., Long, F., Liao, W., Liu, H., (2020). Prediction of anaerobic digestion performance
564 and identification of critical operational parameters using machine learning algorithms.
565 *Bioresource. Technology*. 298, 122495.

- 566 World Biogas Association (WBA). (2021) *Biogas: Pathways to 2030*. Retrieved from
567 worldbiogasassociation.org
- 568 Zhang, L., Loh, K., Zhang, J. (2019). Enhanced biogas production from anaerobic digestion of
569 solid organic wastes: Current status and prospects, *Bioresource Technology Reports*, 5,
570 pp. 280-296.



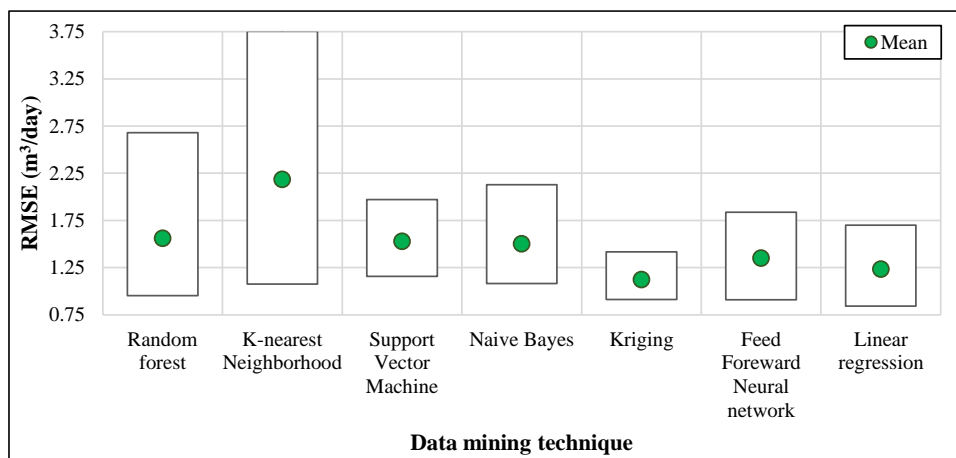
571
572

Figure 1. AI-based framework for the operation of the micro anaerobic digestion plant



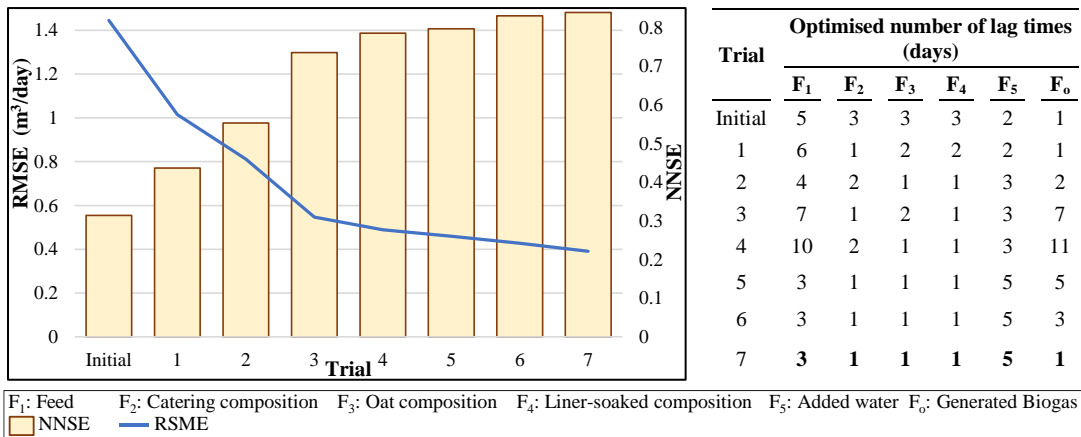
573
574

Figure 2. Schematic diagram of the micro-AD plant used in this study

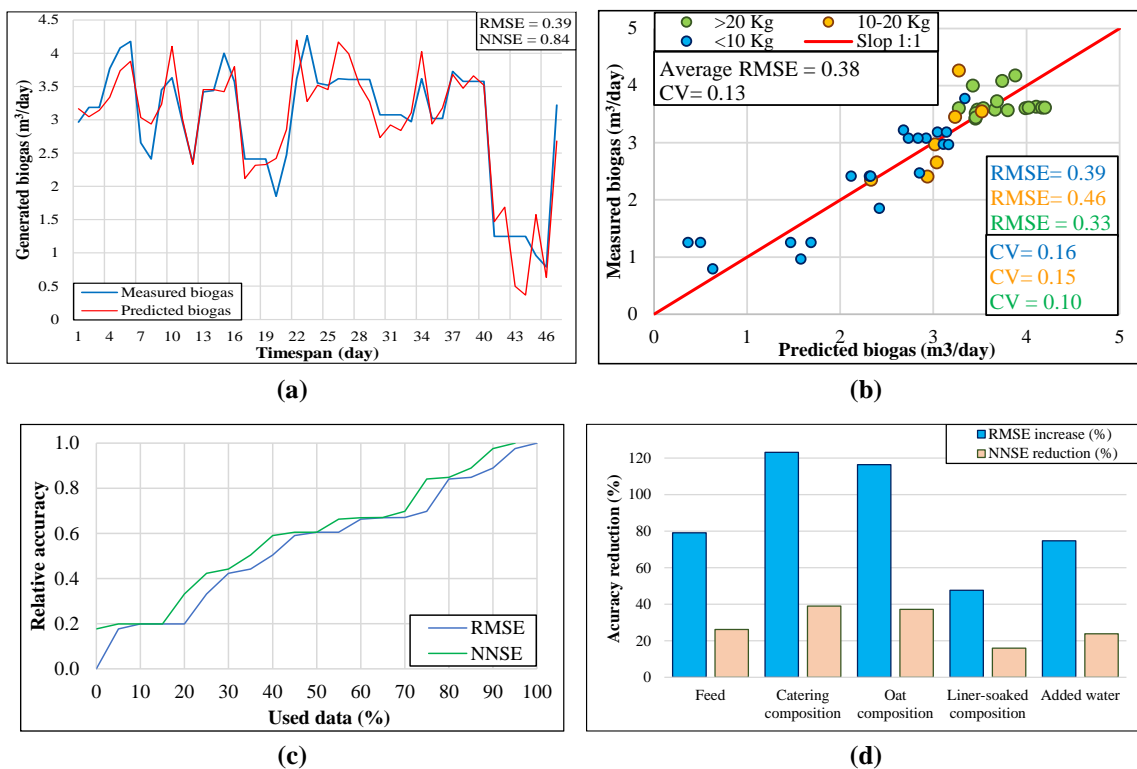


575
576

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



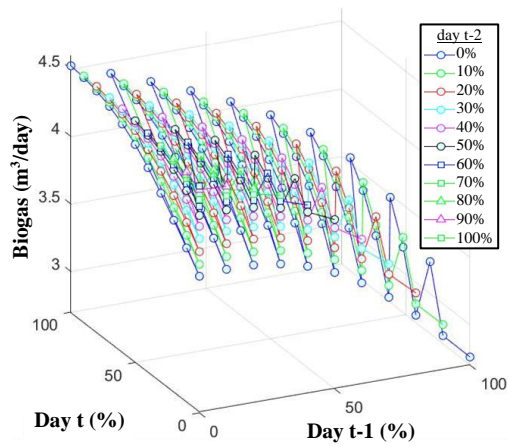
577 **Figure 4. The trend of the SFLA method to specify optimal number of lag times for input variables**



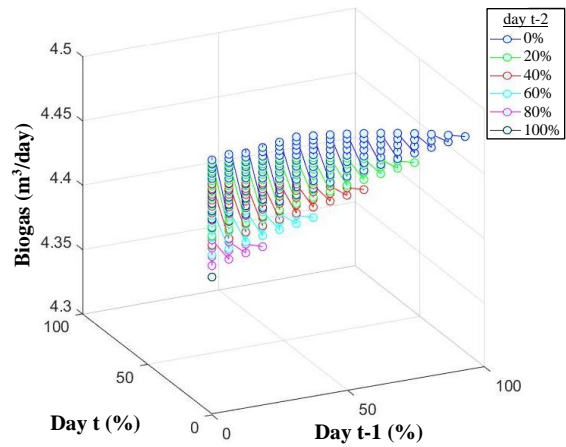
578 **Figure 5. (a) scatter plot of predicted biogas vs corresponding measurements for 1 day ahead, (b)**
 579 **comparison of observations with estimations, (c) relative accuracy based on the percentage of dataset**
 580 **used for model development, (d) Impact of feed compositions in the pre-feed tank on the biogas**
 581 **generation**

582 **Table 1. Optimum condition for the operation of the micro-AD plant for maximum biogas generation**

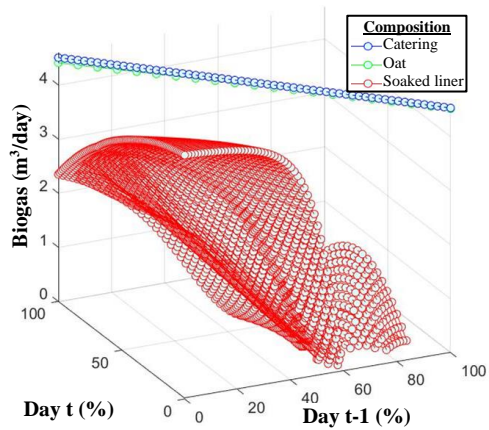
Parameter	Days							
	Predictors (input data)				Predictions			
	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4
Feed	0	0	0	80				
Biogas	-	-	4.11	4.08	4.52	4.23	4.11	4.08
Catering	-	-	55	5				
Oat	-	-	20	0				
Liner	-	-	0	0				
Water	0	0	15	0				



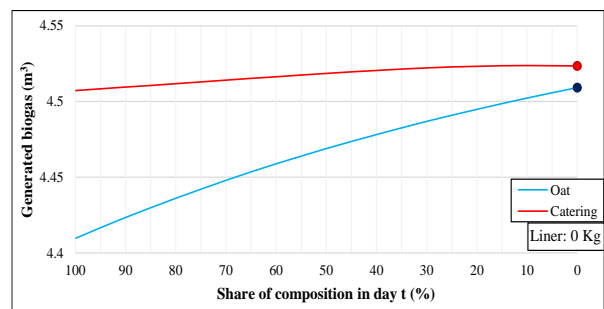
(a)



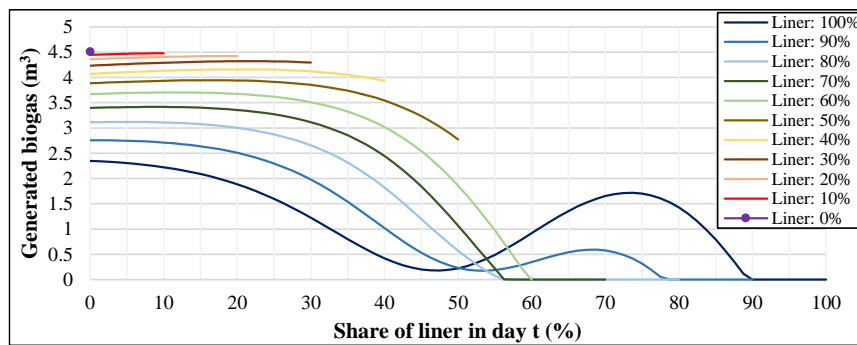
(b)



(c)

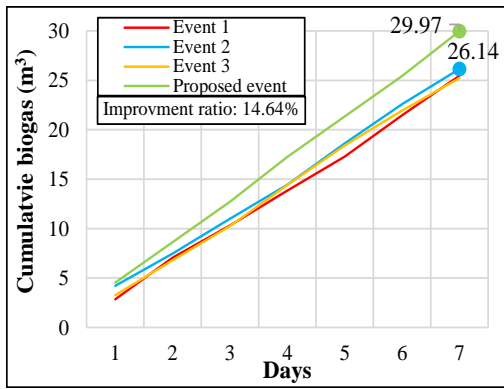


(d)

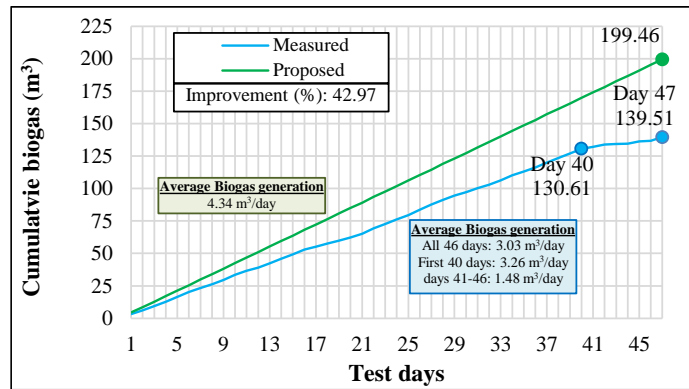


(e)

583 Figure 6. Sensitivity analysis of biogas generation for different decision variables: (a) feed, (b) water, (c)
 584 waste composition; and the impact of the distribution of the optimal values of the pre-feed composition
 585 variables on biogas generation for (d) catering and oat, and (e) liner



(a)



(b)

586
587

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