Çukurova Üniversitesi Mühendislik Fakültesi Dergisi, 39(2), ss. 359-374, Haziran 2024
Cukurova University Journal of the Faculty of Engineering, 39(2), pp. 359-374, June 2024

Artificial Intelligence Predictions of Biomass Power of an Installed Waste Water Treatment Plant

Akın İLHAN[*1] ORCID *0000-0003-3590-5291*

[1]*Ankara Yildirim Bayezit University, Faculty of Engineering and Natural Sciences, Department of Energy Systems Engineering, Ankara, Türkiye*

Geliş tarihi: 12.01.2024 Kabul tarihi: 27.06.2024

Atıf şekli/ How to cite: İLHAN, A., (2024). Artificial Intelligence Predictions of Biomass Power of an Installed Waste Water Treatment Plant. Cukurova University, Journal of the Faculty of Engineering, 39(2), 359-374.

Abstract

In the current study, the power generations obtained from gas turbines of an installed waste water treatment plant were predicted, utilizing artificial intelligence method consisting of artificial neural network (ANN). In this regards, a cumulative of 445 data, found in the power generation data cluster and found in the physical and chemical data clusters has been used in the predictions based on the artificial intelligence association method. Each instant data of these total 445 data corresponds to daily average power generation (P) obtained from gas turbines of the facility and corresponds to physical and chemical parameters including the temperature (T), degree of acidity (pH), conductivity (σ), as well as the daily total volumetric flow of the waste gas to be burned at the gas generator (Q). Accordingly, the best prediction obtained by ANN approach was concluded to generate the statistical accuracy results corresponding to 6.1279% mean absolute percentage error (MAPE), 2.1540 MWh/day root mean square error (RMSE), and 0.9730 correlation coefficient (R) for power generation parameter.

Keywords: Artificial intelligence, Artificial neural network (ANN), Waste water treatment

Biokütle Tipi Kurulu Atık Su Arıtma Tesisinin Enerji Geri Kazanım Analizi

Öz

Bu çalışmada, kurulu bir atık su arıtma tesisinin gaz türbinlerinden elde edilen enerji üretimleri, yapay sinir ağlarından (YSA) yararlanılarak yapay zekâ yöntemi kullanılarak tahmin edilmiştir. Bu doğrultuda, yapay zekâ ilişkilendirme yöntemine dayalı tahminlerde, elektrik üretimi veri kümesinde bulunan ve fiziksel ve kimyasal veri kümelerinde bulunan 445 verinin tamamı kullanılmıştır. Bu toplam 445 verinin her bir anlık verisi, tesisin gaz türbinlerinden elde edilen günlük ortalama elektrik üretimine (P) karşılık gelmekte olup; ayrıca, sıcaklık (T), asitlik derecesi (pH), iletkenlik (σ), ve ilaveten gaz jeneratöründe yakılacak atık gazın günlük toplam hacimsel akışını (Q) da içermektedir. Buna göre, güç üretim parametresi kapsamında, YSA yaklaşımıyla elde edilen en iyi tahminin; %6.1279'luk ortalama mutlak yüzde hata (MAPE) değerine, 2,1540 MWh/gün ortalama karekök hata (RMSE) ve 0,9730 korelasyon katsayısına (R) karşılık gelen istatistiksel doğruluk sonuçlarını ürettiği sonucuna varılmıştır.

Anahtar Kelimeler: Yapay zekâ, Yapay sinir ağları (YSA), Atık su arıtma

[*]Sorumlu yazar (Corresponding Author): Akın İLHAN, *akinilhan@aybu.edu.tr*

1. INTRODUCTION

The increase of fossil fuel powered devices and machines used in many engineering applications has caused increase of greenhouse gas emissions and unfortunately in parallel the global warming in all over the World. Unlike renewable power sources, fossil fuel sources will be depleted soon in the near future. Accordingly, energy production using renewable sources has demonstrated a significant increase in all of the World. As such, nearly whole of the World countries have at least one type of energy plant that operates using a renewable source [1]. Therefore, especially in the last decades, energy plants powered by renewable sources have an exceeding spreading potential against the energy plants powered by conventional fossil fuels. On the other hand, energy generation obtained from biomass sources is among these types, and biomass is one of the most substantial renewable type today and it is utilized in energy production in many countries of the World.

In this regards, the share of the World total installed biomass power is exhibited in Figure 1, presented as annual distributions starting from 2010 to present day, and as well its 2025 projections are also included. This figure also provides a projection of the other energy sources in the near future, i.e., by 2025. It is indicated in this figure that reaching 2025, installations of the fossil fuels will significantly decline back in terms of the percentages. Even hydropower, which is one of the leading types of renewable energy and which has been applied for the first time in the World as a renewable source and has a very large amount of installation capacity considering all over the World, has decreased considerably in percent today. Today, the power generation capacities belonging to hydraulic sources and solar sources are almost equalized. Besides, by 2025, it is obvious that other renewable energy types, especially solar and wind energies, will result hydraulic energy installations to decline further in percentage due to the increase of those in installations all around the World. As indicated in Figure 1, although bio-power constitutes a low amount in total percentage; the energy generation potential of plant and animal waste, which is the source of bio power, is quite high and only a small amount can be converted into useful energy yet. This situation will make bio power a very important and widely used fuel type in the future.

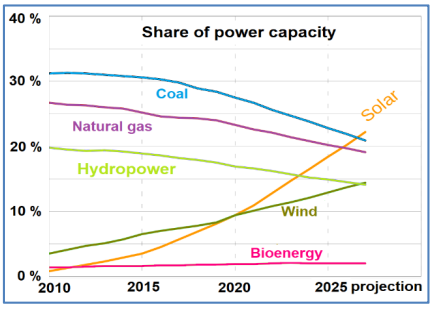

Figure 1. 2025 projections of the renewables compared to fossil energy sources

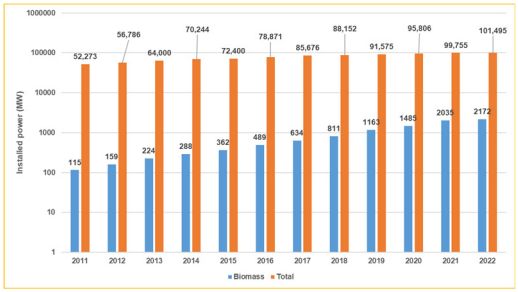

Figure 2. The biomass and total power installations of Türkiye [2]

This situation manifests itself in this way, in Türkiye as well. Renewable power plant installations also in Türkiye, have accelerated in the last few decades. However, the biomass installations of the country also boosted around 20 times in the last twelve years. As presented in Figure 2, while total installed power of the country has increased twice in the considered time range, reaching 101,495 MW at the end of 2022, from 52,273 MW by 2011; whereas, the biomass installations of the country have increased approximately 20 times in the same year range, reaching 2,172 MW at the end of 2022, from only 115 MW by 2011 [2].

The scientific studies taking part in the literature include estimation implementations of physical and engineering problems conducted mainly in three different ways. Those consist of physical ways,

statistical approaches, and machine learning or artificial intelligence tools [3,4]. Physical methods are also referred as the plain methods and those use some physical information. The physical information could be topographic data, roughness, temperature, pressure, and even obstacles [5]. Besides, the certainties of the statistical approaches are rather limited. In the study of Liu et al. [6], the developed autoregressive integrated moving average model was suggested in order to obtain predictions of wind speed, considered in the railway strong wind warning system. Besides, future wind speed values that have not yet realized, i.e., one day later's or two day later's wind speed data of a chosen location of North Dakota were predicted using fraction-ARIMA, which is again a type of statistical models [7].

On the other hand, recent novel techniques including ANN (artificial neural network) as well as different machine learning methods are also implemented in data forecasting. Especially, ANN and machine learnings have been widely applied to predictions of wind speed, until now [3]. Also, solar energy being a more recent technology fairly utilizes predictions performed by machine learning. In this regards, Khosravi et. al. [8] have implemented an algorithm of machine learning for hourly solar irradiance estimations of the selected region. Besides, LSTM approach was exerted on predictions of the experimental results of wind power, in the study of Shi et. al. [9]. The LSTM estimation results had presented sufficient accuracy compared against the measured data [9].

Waste-water is formed of clean water which was used by people and later became to be dirty. As soon as it becomes dirty, it may consist of many undesired substances such as bacteria, food waste, chemicals, and maybe some other undesired substances. For instance, at every place of the houses, waste water could be produced. These locations may include the sinks, dishwashers, and washing machines. In this context, the target of obtaining treatment of the waste-water is to obtain removal of undesired biological or chemical substances dissolved inside the water as much as possible [10].

Exclusively, in the last decades, data predictions have become significantly important. Among those, the estimations of the future atmospheric conditions can be given as an example. This is actualized in order to determine future values of air temperature or pressure, the wind speed and direction, the precipitation conditions, else the humidity. Those are rather helpful, for instance, in order to take measures against natural disaster including floods or hurricanes. During the travel of airplanes or road vehicles, the information related with the precipitation data or air temperature are certainly required. Besides, to obtain safe travels of ferries or ships in the sea, ocean, or rivers; predictions performed on the wind speed, water wave height, or water flow speed are generally utilized. In this study, as an innovation, the prediction of the data consisting of the gas generator daily power output (P), with regards to the gas generator daily gas consumption (Q), wastewater temperature (T), wastewater conductivity (σ) and the degree of the acidity (pH) of waste-water of the past time-series of an established wastewater treatment plant, was executed. Although this study is not a hydrological study; it includes the estimation of the energy production obtained in gas turbines by the combustion of methane gas, which is formed and emerged as a result of hydrological wastewater streams. On the other hand, data estimations have the same general principles, as in the case of predicting hydraulic data. In this regards, an establishment of 7 different models was performed, and in this way, the aforementioned physical and chemical properties of the past wastewater data and the daily power output (P) data are functionally related. Accordingly, a total of 7 main different models was tried and the best model was demonstrated, in obtaining best predictions of the power generation (P) obtained from the combustion of the methane gas of the waste-water. In this way, the estimation quality was compared between these seven main models formed of inputs including the historical time series of the gas generator power output (P), compared against the inputs including the physical and chemical properties of the waste-water.

2. MATERIALS AND METHODS

2.1. ANN

Artificial neural networks are mainly and widely used in computational processes, and they resemble a network of nerve cell of biological systems. The neural network is a recently utilized computer architecture. It is a novel algorithmic structure and has a rapid and a superior learning capability, compared to conventional computers or calculators. It provides some basic computational operations to easily solve or numerically converge to a solution of complex or nonlinear problems, that are usually impossible to be solved by analytical means [11, 12]. A neuron is the basic processing element of a neural network. The back-propagation (BP) algorithm necessarily starts with the calculation of the output layer, which is the only layer where the desired outputs are available but no intermediate layer outputs, as demonstrated below in Figure 3 [13];

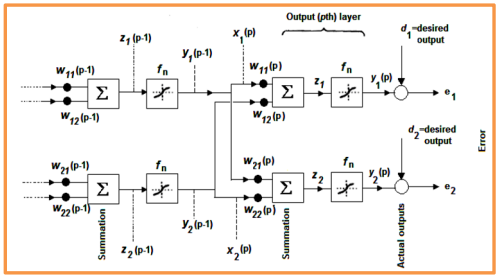

Figure 3. A perceptron of multilayer

2.2 Statistical Error Analysis

The quality test of the artificial intelligence estimations has been executed using three different statistical error parameters. As required, the predicted outcomes are compared with the real measured or observed data, using the mean absolute percentage error (MAPE), the correlation coefficient (R), and the root mean square error (RMSE). As the name implies, depending on the prediction accuracy, MAPE can take values in between 0% and 100%. Besides, RMSE can take any values in the positive real data set. However, the approach of MAPE and RMSE close to zero indicate higher accuracy and quality of the computations. Finally, the correlation coefficient can take any real value in the data range of $0 \leq R \leq 1$, else also including the boundary integers of 0 and 1. However, while the correlation coefficient (R) approaching to zero exhibits lower correlations between the real data cluster and the predicted data cloud, whereas, the correlation coefficient (R) approaching to unity (one) shows higher correlations between two data sets.

Following three equations give the mathematical definitions of MAPE, RMSE, else R, respectively [14]:

$$MAPE = 100x \frac{\sum_{i=1}^{N} \frac{|o(i)-p(i)|}{o(i)}}{N} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[p(i)-o(i)]^2} \quad (2)$$

$$R = \frac{\left(\sum_{i=1}^{N}[p(i)-\bar{p}][o(i)-\bar{o}]\right)}{\left(\sqrt{\sum_{i=1}^{N}[p(i)-\bar{p}]^2}\sqrt{\sum_{i=1}^{N}[o(i)-\bar{o}]^2}\right)} \quad (3)$$

In Equations (1), (2), and (3), i is used as the data sequence in the data cloud to call any instantaneous data. Again in these three equations, N corresponds to the cumulative amount of the members of all data set. Besides, the functions $o(i)$ and $p(i)$ are utilized to call any instantaneous data, respectively in the observed data cloud and the predicted data cloud, in these three equations. Multiplying the error by 100 in Equation (1) makes the error value obtained in percentage. In Equation (3), the abbreviation $\bar{o}$ indicates the average value of all real observed data called between i=1 to i=N found in the real observed data cloud. Finally, the denotation $\bar{p}$ shows the mean value of whole predicted or simulated data called between i=1 to i=N found in the predicted data cloud [14].

3. RESULTS AND DISCUSSIONS

Machine learning or artificial intelligence data estimation finds a lot of use in both physical and engineering applications. In this way, unrealized future value of the physical data that is in

consideration could be predicted with low margins of errors, using machine learning, in the case of having or forming a continuous data function. The advantage of such types of modelling involves the input data to be only the historical past time series of the considered physical parameter, or some other physical input parameters, without any detailed technical knowledge of the measurement device that is used or without performing to find solutions to complex mathematical problems.

This study was executed using a total of 445 daily average physical and chemical type of data of an installed waste water treatment plant. The data consisted of power generation (P) data of gas turbines of installed waste water treatment plant as well as the physical and chemical properties of waste water of water treatment plant. As such, the physical parameters include the temperature (T) and the conductivity of wastewater (σ), whereas, the chemical properties include the degree of the acidity of the wastewater (pH); besides, the wastewater was evaluated in the wastewater treatment plant for treatment processes as well as energy generation in the gas generators. The energy generation from the gas turbines is obtained by the generated methane gas due to the decay of the biomass. Namely, the methane (CH_4) and stabilized biomass are generated by the biomass digestions [15,16].

Among the total of 445 daily average data; 80% of the cumulative data was utilized in order to obtain proper trainings of the algorithms, whereas the rest of 20% was used in testing of the algorithms. Accordingly, in the simulations of the artificial intelligence, ANN tool was used. Figure 4 illustrates all temperature (T) data cloud allocated for training and testing, shown in orange and purple colours, respectively. As shown in this figure, the temperature (T) of the waste water is in the range of $12.400\ ^oC \leq T \leq 26.820\ ^oC$. Else, the data analysis has demonstrated that mean value of the waste-water temperature is 21.061ºC, in this data range.

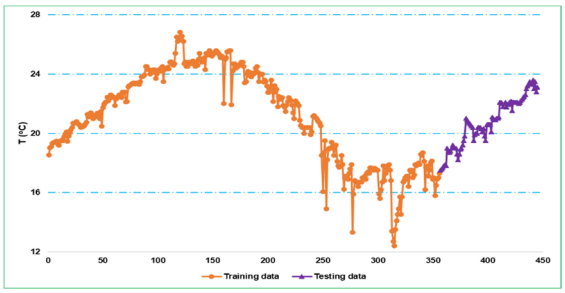

Figure 4. The training and testing data clouds of temperature (T) parameter

Figure 5 exhibits the entire electrical conductivity (σ) data cloud separated for training and testing, shown respectively with grey and yellow colours. In this regards, the conductivity (σ) of the waste water is in the range of $540.000\ \mu S/cm \leq \sigma \leq 1{,}741.000\ \mu S/cm$. The waste-water conductivity (σ) data analysis had revealed that the average value of data in this range is $1{,}315.250\ \mu S/cm$.

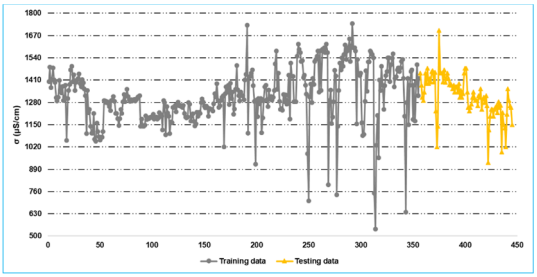

Figure 5. The training and testing data clouds of conductivity (σ) parameter

The mean value of the degree of the acidity (pH) was observed to correspond 7.493 in the data range of $6.410 \leq pH \leq 8.570$. Since mean value of the degree of the acidity (pH) is over the value of 7.00, the analysis has revealed that the general structure of the solution is alkaline. It is also already observed from Figure 6 that instantaneous values of the measured pH data are generally over pH=7 line, implying this alkalinity chemical structure. Ultimately, Figure 6 indicates whole degree of acidity (pH) data cluster discretized for training and testing presented in blue and green colours, respectively.

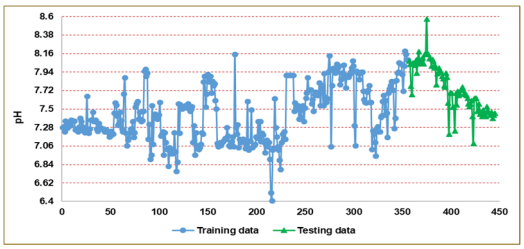

Figure 6. The training and testing data clouds of degree of acidity (pH) parameter

The average value of the daily average values of gas generator gas consumptions (Q) was determined to be 11,767.110 m^3. Besides, the gas consumption (Q) data cluster included values in the range of $0.000\ m^3 \leq Q \leq 18{,}832.000\ m^3$. Accordingly, Figure 7 presents the value distributions belonging to whole gas generator daily gas consumptions (Q), in which this cluster is similarly discretized for training and testing presented in brown and red colours, respectively. This figure involves only the quantity of the consumed methane gas at the gas turbines, to generate electricity, not all of the quantity of the generated gas due to the decay of biomass forming methane gas. In reality, the total methane gas amount produced at the wastewater treatment plant is greater than the total amount of methane gas burned in the gas turbines, given in this figure as daily distributions. However, some of them escape or being lost from the facility without being burned in the gas turbines. Under these conditions, the total amount of methane gas burned in the gas turbines to generate electricity is reported to be 5.236 hm^3, in this 445 days' time interval.

Finally, the analysis of the data has exhibited that the data range for the gas generator daily total energy generation (P) to be in the range of $0\ MWh \leq P \leq 48.780\ MWh$. Else, the medium value of the power generation (P) was shown to be 31.911 MWh. All in all, Figure 8 demonstrates whole gas generator daily total energy generation (P) data cluster split for training and testing displayed light grey and purple colours, respectively. In gas generators, this energy produced by the combustion of methane gas can be considered as recovered energy. Because the facility needs a serious amount of energy consumption to carry out the wastewater treatment activity. This energy obtained from burning of methane gas in the gas generators, meets a significant part of the energy required for the facility to continue its function, and as a result, it is possible to significantly reduce the energy drawn from the network or produced by diesel generators in case of power outage. Because the energy drawn from the electricity network or the diesel fuel used for burning in diesel generators constitute a serious expense item. In this way, the expense items constituting high costs would be greatly diminished. Besides, in the analysed 445 days, it is reported and shown that total energy generations by burning of the methane gas in the gas turbines correspond to 14.200 GWh of power.

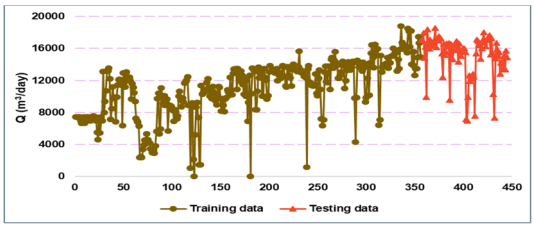

Figure 7. The training and testing data clouds of the gas generator daily gas consumption (Q)

Figure 8. The training and testing data clouds of the gas generator daily total energy generation (P)

Regarding the predictions of the generated power (P) of the facility that is executed by the artificial neural network (ANN), in this context, the daily total energy generation (P) of the gas turbine is defined according to the mathematical exponential functional relation with respect to the waste water temperature (T), the degree of acidity of the waste water (pH), the conductivity of the waste water (σ), as well as the daily total volumetric flow of the waste gas to be burned at the gas generator (Q).

Among these 4 physical and chemical parameters that were used as inputs; the waste water temperature (T) was considered in "Kelvins" to construct a better functional relation between inputs and output, the degree of the acidity (pH) was considered as usual as a dimensionless quantity to be in between 0 and 14, the conductivity (σ) was handled in $\mu S/\mu m$, whereas gas generator daily gas consumption (Q) was taken with the unit of m^3/day. Finally, gas generator daily total power generation (P) which is the dependent functional parameter was obtained at the unit of MWh/day. But, the instantaneous data points in Figures 4-8 were provided in "degree Celcius", $\mu S/cm$, non-dimensional, m^3/day, and MWh/day, respectively for T, σ, pH, Q, and P. Accordingly, two units of "degree Celcius" and $\mu S/cm$ were considered slightly different in ANN computations compared to the demonstrated ones, respectively in Figures 4 and 5 for T and σ. Namely, the instantaneous "degree Celcius" temperature data were simply converted to Kelvin by adding the value of 273.15 to each instantaneous data, and the instantaneous conductivity $\mu S/cm$ data were simply converted to $\mu S/\mu m$ by dividing each instantaneous data by the value of 10^4. In this way, better quality algorithms were created between inputs and outputs. Shortly, the average daily values of the parameters such as temperature (T), the degree of acidity (pH), the conductivity (σ), as well as the daily total flow of the methane gas (Q) were either used directly or modified with respect to the formerly defined ranges of these physical and chemical parameters; and those were finally processed with the exponential function to obtain the correspondence output power generation (P) of the gas generator of the facility.

The ANN algorithm was structured at the epoch number of 500 and the goal of the training parameters was adjusted to $5x10^{-19}$, implying that the iterations would stop as soon as it converges below this numerical value. In ANN computations, as similar to computations of historical time-series, for 4 inputs and 1 output; a total of 356 waste water data was considered as averages or cumulative in training stage, whereas, a total of 89 data was considered as averages or cumulative in the testing stage.

The schematic representation of configured structures of ANN models is exhibited in Figure 9. Accordingly, a total of 7 main ANN models were designed and explained in this figure. Besides, the names of these main models are indicated in this figure, with green color. The desired output in all models is the output power (P) of the gas turbine, in which the gas turbine generates power by consuming methane gas that was obtained by the decay of the ingredients of the waste-water. This situation is depicted by P designation in these 7 main models, shown in Figure 9. Actually, power generation (P) forms the output of these models. Besides, two bias are required in each main model to relate the inputs and outputs with a proper equation. Those are shown as Bias 1 defined at the hidden layer and Bias 2 defined at the output layer, and indicated in this figure in this direction. At each main model, as presented in this figure, a total of 8 different sub-models were created. Accordingly, the hidden layer (HL) number was chosen in the range of $3 \leq HL \leq 10$, with an increment of $\Delta HL = 1$, between each consecutive sub-models, i.e., constituting a total of 8 different sub-models, for each main model. To describe the inputs for the main models; initially, Model 1 was generated according to the inputs of waste-water temperature (T) and the daily gas consumption of the methane gas (Q). Secondly, Model 2 was formed again using two inputs, however this time, the degree of the acidity (pH) and the daily gas consumption of the methane gas (Q). This was followed by Model 3, including the inputs of the electrical conductivity of the waste-water (σ) as well as the daily gas consumption of the methane gas (Q). Up to Model 3 and including this model, only two inputs were considered as indicated in Figure 9. But, between Model 4 and Model 6, three inputs were composed to relate the inputs with the output parameter of the power generation (P). In this regards, in Model 4, the temperature of the waste-water (T), the degree of the acidity (pH), and the daily gas consumption of the methane gas (Q) were considered as inputs of ANN algorithm. Whereas, in Model 5, the temperature of the waste-water (T), the electrical conductivity of the waste-water (σ), and the daily gas consumption of the methane gas (Q) were selected as inputs of the output power generation (P). Model 6, the final build with 3 inputs consists

of the degree of the acidity (*pH*), the electrical conductivity (σ), and the daily gas consumption of the methane gas (Q). Eventually, Model 7 is the single model formed of four inputs. In this context, it is designed to imply the temperature of the wastewater (T), the degree of the acidity (*pH*), the electrical conductivity (σ), and the daily gas consumption of the methane gas (Q), in order to functionally relate these inputs with the output parameter of power generation (P). All in all, 8 models were created for each main model, that have been obtained by the alteration of *HL* number in the range of $3 \leq HL \leq 10$; that resulted a total of 56 models for 7 main models and the input to output simulations were performed accordingly. All of the models include the forecasts at training and testing stages together. Accordingly, total of 445 data has been shown together with its corresponding predictions in these models.

In this study; the ANN artificial intelligence algorithm was developed using daily gas consumption of the methane gas (Q), the acidity level of the wastewater (*pH*), the electrical conductivity of the wastewater (σ) and the temperature values of the wastewater (T). In physics, a considered type of energy per unit mass has a certain amount of energy content. For example, air blowing as wind at a certain amount of speed has a certain energy content per unit kg, whereas water flowing as river again at a certain amount of speed has a certain energy content per unit kg. If their units are expressed as J/kg, and if these values of the energy content are multiplied by the amount of the substances belonging to the relevant energy type displaced per unit time, the amount of the energy rate at a certain mass flow rate from the expression of (J/kg).(kg/s) will be obtained. In other words, the energy produced in physics has an inseparable connection with the substance flow rate. This also applies to methane gas entering to the gas generator. In other words, the methane gas energy produced from the gas generator has a directly proportional relationship with the methane gas flow rate entering to the gas generator. For this reason, in all 56 tested models in total, that are obtained consisting of built 7 main models and 8 intermediate models created for each main model, the parameter of daily gas consumption of the methane gas (Q) was used continuously, without any change, considered in training, testing stages as well as during the creation of the ANN algorithm. As a result, the predictions of the energy production of methane gas formed from the wastewater have been performed by considering three other parameters such as wastewater temperature (T), acidity level of wastewater (*pH*), and electrical conductivity of wastewater (σ), with the methodology of selecting any of those three alone or by randomly selecting two of the three, or by considering all three together; however, at all conditions always using the daily gas consumption of the methane gas (Q).

Figure 10 exhibits the results of main Model 1, formed of two inputs, including the output of the real measured data of gas generator daily total power generation (P), and its counterpart output ANN predictions obtained at *HL*=3, *HL*=4, *HL*=5, *HL*=6, *HL*=7, *HL*=8, *HL*=9, and *HL*=10. The selected input parameters are both physical ones including temperature (T) and the daily consumption of the methane gas (Q). Therefore, this main Model 1 includes the results of a total of 8 sub-models, shown at the same figure. In this figure, the light blue, pink, black, yellow, red, green, dark blue, brown and grey colors were used to indicate functional distribution curves of real power (P) data as well as its estimations obtained at *HL*=3, *HL*=4, *HL*=5, *HL*=6, *HL*=7, *HL*=8, *HL*=9, and *HL*=10, respectively.

On the other hand, Figures 11 and 12 show the output power generation (P) results of main Models 2 and 3, formed of two inputs, again considering same ranges of hidden layers (*HL*). In this regards, one input parameter is chemical, while the other is physical in main Model 2; respectively corresponding to the degree of the acidity (*pH*) and the daily consumption of the methane gas (Q). On the other hand, both parameters are physical in main Model 3. I.e., the electrical conductivity (σ) as well as the daily consumption of the methane gas (Q) were selected as inputs to obtain output parameter of power generation (P). Same logic of coloring was also taken into account in both figures for the functional curves belonging to real measured data

and corresponding estimates at different hidden layers (*HL*). Similarly, Figures 11 and 12, each involve the results of 8 sub-models obtained for main Models 2 and 3.

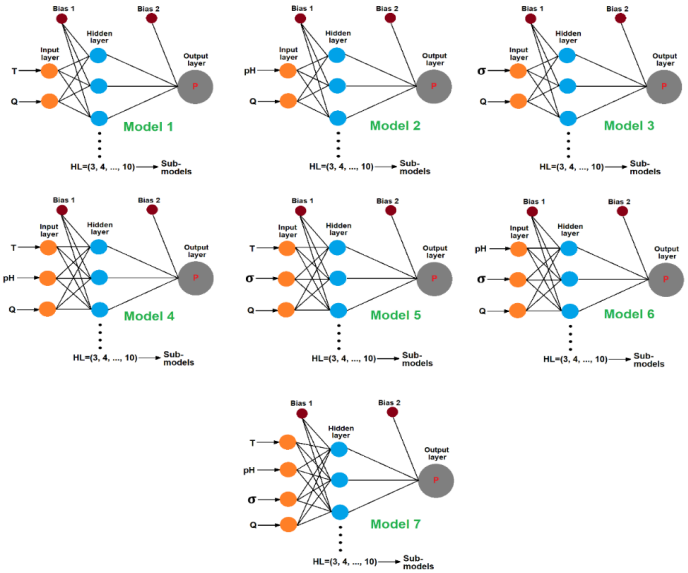

Figure 9. Seven main models of ANN structure

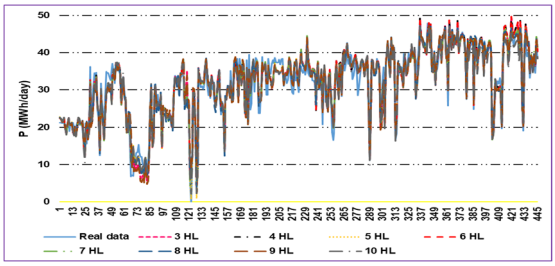

Figure 10. Model 1 results for gas generator daily total energy generation (*P*)

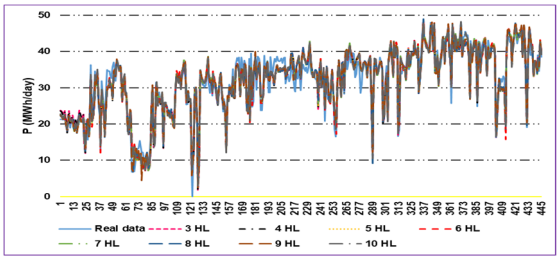

Figure 11. Model 2 results for gas generator daily total energy generation (*P*)

Figure 13 presents main Model 4 data distributions, now formed of three inputs, including the real output power (*P*) data function given with respect to the predictions obtained considering different hidden layers (*HL*) studied in the range of $3 \leq HL \leq 10$. The inputs are formed of two physical and one chemical parameter. While the degree of the acidity (*pH*) was chosen as the chemical parameter, whereas, the temperature (*T*) as well as the daily consumption of the methane gas (*Q*) were selected as the physical parameters. Again 8 sub-models are included in this figure as well as the real data function is included. As similar to former main models, the same colors and patterns have been taken into account to indicate the real data function and the estimations obtained at different hidden layer (*HL*) numbers.

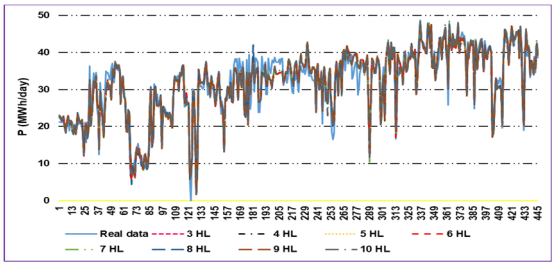

Figure 12. Model 3 results for gas generator daily total energy generation (*P*)

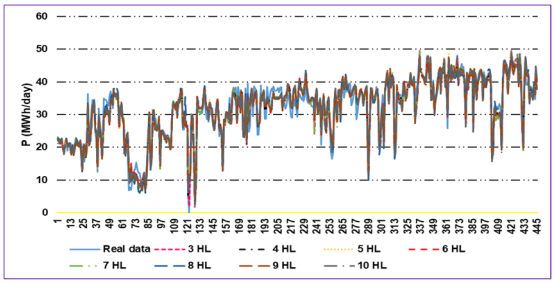

Figure 13. Model 4 results for gas generator daily total energy generation (P)

The data distributions of main Model 5 consisting of real output power (P) data set versus the predictions at different hidden layer numbers, formed of again three inputs, are exhibited in Figure 14. All of the studied parameters in this main model are physical parameters, and no chemical parameter was considered in this main model. In this regards, the temperature (T), the electrical conductivity (σ), as well as the daily consumption of the methane gas (Q) are used as physical parameters of the inputs to obtain the output of power generation (P). Again comparison of 8 sub-models has been revealed in this figure. Same curve colors and curve patterns were utilized for different hidden layers (HL), considered in the range of $3 \leq HL \leq 10$, as similar to previous four main models.

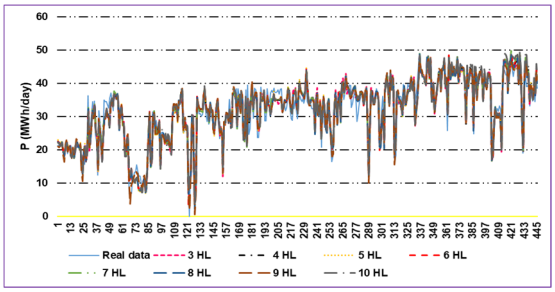

Figure 14. Model 5 results for gas generator daily total energy generation (P)

Figure 15 displays results of main Model 6 involving real output power (P) data cloud given according to the estimations obtained at 8 distinct hidden layer numbers. Model 6 was constituted of three inputs to obtain power generation (P) which is the output parameter. Apart from the chemical parameter of the degree of the acidity (pH); the electrical conductivity (σ), and the daily consumption of the methane gas (Q) were selected as the physical parameters. In main Model 6, the study of 8 different hidden layer numbers resulted again the comparison of the estimated power generation results of 8 different sub-models that were compared to the real measured output power (P). In parallel to the former implemented main models, same curve colors and curve patterns were used in this main Model 6 for different hidden layers (HL), considered in the range of $3 \leq HL \leq 10$.

Final main model, i.e., Model 7 is the only one formed of 4 inputs, in which one is the chemical, the rest are the physical parameters. Eventually, the outcomes of this model are demonstrated in Figure 16. In main Model 7, the degree of the acidity (pH) was used as chemical parameter input, whereas, the temperature (T), the electrical conductivity (σ) and the daily consumption of the methane gas (Q) were utilized as the physical parameter inputs. In this way, the advantages or disadvantages of using all four of the inputs were tested in main Model 7, according to the incomplete use of some of the inputs which were tested in the former main models. On the other hand, as similar to the former main models; again the outcomes of 8 different hidden layer (HL) numbers were studied and compared with respect to the real measured data cloud of the gas generator daily total energy generation (P). To indicate the real or estimated data cloud function curves, both same colors and same patterns of the previous main models were taken into account as denoted in Figure 16.

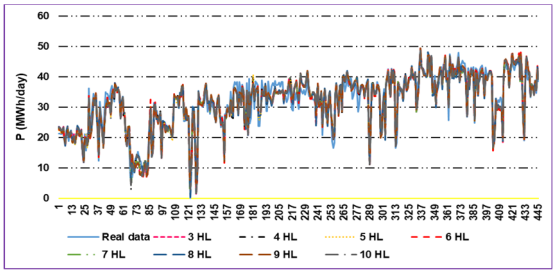

Figure 15. Model 6 results for gas generator daily total energy generation (P)

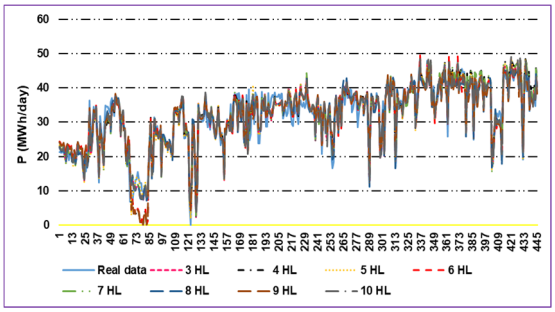

Figure 16. Model 7 results for gas generator daily total energy generation (*P*)

The statistical accuracy results of 7 main models as well as 56 sub-models have been presented in Tables 1 and 2, including MAPE, RMSE, and *R*. In this regards, while Table 1 gives the statistical errors of main Models 1, 2, 3, and 4; whereas, Table 2 provides the statistical errors of main Models 5, 6, and 7. In both tables, the best results for each main model giving the superior statistical accuracy outcomes are indicated by bold color to be distinguished from other sub-models provided at different hidden layer numbers (*HL*). Besides, the best result of all 7 main models is exhibited with both bold and italic patterns. In this context, MAPE, RMSE, and *R* results of main Model 5 at 9 *HL* found in Table 2 generates the best outcomes, respectively corresponding to 6.1279%, 2.1540 MWh/day, and 0.9730. The estimation function of this result is reported as the function that converges most to the real observed values of the gas generator daily total energy generation (*P*), among 56 ANN models tested in this ANN study.

Apart from the best result of all ANN computations, the simulation performed at 5 *HL* of main Model 1 found in Table 1, the computation executed at 9 *HL* of main Model 2 found in Table 1, the study actualized at 6 *HL* of main Model 3 found in Table 1, as well as the trial conducted at 10 *HL* of main Model 4 found in Table 1, performed better results than their counterparts found in the same main models. Similarly, in Table 2, sub-model results at 9 *HL* of main Model 5, sub-model outcomes at 8 *HL* of main Model 6, else statistical error results of 5 *HL* of main Model 7 have generated relatively preferable results than their counterparts found in the same main models.

Although, the estimated data function distribution belonging to the sub-model of 9 *HL* of main Model 5 was shown in Figure 14 with brown color, since this result outcame with the best convergence to the real observed values of the gas generator daily total energy generation (*P*); for this reason, to obtain a better visualization of the prediction quality of this model, the data distributions of the actual function and the predicting function at this model are displayed in Figure 17. In this figure, while the real observed value distributions are indicated with a curve of continuous orange color, whereas, its corresponding estimations performed at 9 *HL* of main Model 5 are demonstrated with dashed type curve of light blue color. When the distributions of the actual production data given in Figure 17 are compared with the forecast distributions that provide the best estimate and obtained in 9 *HL* of main Model 5; it is seen that two function curves overlap each other quite well and the estimated curve captures the real production curve well, even including the sudden break points. The correlation coefficient of 0.9730 *R* for this superior model also indicates the high compatibility of this implemented model. On the other hand, the real measured data set was indicated in Figure 8, to be in the range of $0\ MWh \leq P \leq 48.780\ MWh$. Besides, the forecasting data set of the model performed at 9 *HL* of main Model 5, had generated the predicting range to be in between $0.544\ MWh \leq P \leq 47.430\ MWh$. In this context, the discrepancy errors of the boundary values of this predicting range with respect to the real data cluster are reported to be significantly small.

Although it is thought that using more input parameters will strengthen the relationship with the output, and accordingly, generate better prediction results when associating a physical parameter with mathematical equations by connecting it to other physical parameters utilizing the artificial intelligence method, this is not always the true. During the introduction of an additional parameter to the artificial intelligence while ensuring the correlation of the relationship provides sometimes higher success, this is true up to a certain threshold value. Namely, introducing additional parameters to artificial intelligence above the threshold value,

may reduce the success of the algorithm. When the studied main models are evaluated, during together utilization situation of the parameters consist of wastewater temperature (T), wastewater electrical conductivity (σ) and daily gas consumption of the methane gas (Q) taking part in the framework of the 5th main model; in addition to these in the 7th main model, adding the acidity level (pH) parameter of wastewater to the artificial intelligence, does not improve the modeling, on the contrary, it worsens it. In this case, it is observed that the acidity degree (pH) parameter of the wastewater, in addition to the electrical conductivity (σ) parameter, does not actually provide a further prediction improvement, on the contrary, it relatively worsens the prediction error parameters. This is due to the fact that the electrical conductivity (σ) parameter and the wastewater acidity (pH) parameter, all in all, are directly related to each other in a physical aspect. The more electrical conductivity (σ) occurs in the wastewater environment, the further the wastewater gets away from its pH neutral environment and more increase will occur in the acidity level (pH) of the wastewater in the acidic or basic direction. For this reason, during algorithm training with the artificial intelligence, as in Model 7, simultaneous utilization of two parameters consisting of electrical conductivity (σ) and the degree of acidity (pH) along with the other two parameters consisting of temperature of the waste-water (T) and the daily consumption of the methane gas (Q); does not further improve the mathematical learning between the instantaneous values of the artificial intelligence's data cloud, conversely worsens it.

Table 1. Statistical accuracy results of ANN computations for Main Models 1, 2, 3, and 4

Main model number	Sub-model (*HL*)	MAPE (%)	RMSE (MWh/day)	R
1	3	7.5360	2.4845	0.9642
	4	6.8609	2.3955	0.9664
	5	**6.7281**	**2.2849**	**0.9696**
	6	6.8729	2.3934	0.9665
	7	6.6590	2.3414	0.9678
	8	6.8875	2.3453	0.9673
	9	7.9915	2.5423	0.9615
	10	6.7889	2.2590	0.9689
2	3	6.9177	2.3609	0.9661
	4	6.9086	2.3718	0.9658
	5	6.9107	2.3743	0.9657
	6	6.6623	2.2931	0.9682
	7	6.7451	2.3305	0.9671
	8	6.4806	2.3313	0.9670
	9	**6.1985**	**2.2516**	**0.9695**
	10	6.5529	2.2445	0.9695
3	3	6.4294	2.3052	0.9677
	4	6.4940	2.3098	0.9676
	5	6.4935	2.3119	0.9675
	6	**6.1689**	**2.2574**	**0.9691**
	7	6.3413	2.2815	0.9687
	8	6.4656	2.3025	0.9680
	9	6.4177	2.3050	0.9677
	10	6.3311	2.2704	0.9689
4	3	6.7890	2.3103	0.9681
	4	6.6291	2.3197	0.9677
	5	6.8251	2.3460	0.9671
	6	6.7508	2.3487	0.9667
	7	6.7897	2.3346	0.9677
	8	7.0165	2.4263	0.9651
	9	6.7995	2.3430	0.9667
	10	**6.3995**	**2.2848**	**0.9693**

Table 2. Statistical accuracy results of ANN computations for Main Models 5, 6, and 7

Main model number	Sub-model (*HL*)	MAPE (%)	RMSE (MWh/day)	R
5	3	6.9356	2.4015	0.9656
	4	6.9742	2.4134	0.9665
	5	7.1193	2.4268	0.9647
	6	6.7679	2.3788	0.9670
	7	6.7910	2.3468	0.9682
	8	6.7640	2.3550	0.9672
	9	***6.1279***	***2.1540***	***0.9730***
	10	7.2975	2.5900	0.9633
6	3	6.9145	2.3581	0.9663
	4	6.6185	2.2525	0.9693
	5	6.5877	2.2697	0.9689
	6	6.8438	2.3049	0.9680
	7	6.4537	2.3227	0.9673
	8	**6.5098**	**2.2350**	**0.9697**
	9	6.7482	2.3250	0.9674
	10	6.5699	2.2501	0.9694
7	3	6.8162	2.3623	0.9665
	4	9.0479	2.9186	0.9550
	5	**6.4265**	**2.2466**	**0.9706**
	6	8.9553	2.8297	0.9532
	7	6.8448	2.3971	0.9670
	8	7.1917	2.4647	0.9628
	9	8.9645	2.8446	0.9528
	10	6.8151	2.3540	0.9668

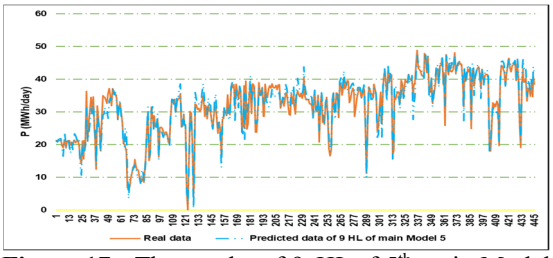

Figure 17. The results of 9 *HL* of 5th main Model for gas generator daily total energy generation (*P*)

Artificial neural network (ANN) type of artificial intelligence (AI) on the data prediction is related with the physical algorithmic interpretation of input data to obtain corresponding magnitudes of the output data with an acceptable error range. Here, the amount and the type of selected input parameter and even their physical units are important in capturing the relationship between input and output variables. On the other hand, it also becomes important as the number of the hidden layers (HL) selected will increase and decrease the mathematical relationships in the equation that captures the output value in the mathematical modeling. In the study, appropriate mathematical models were created by selecting the input parameters in different combinations, that is, by creating different main models; during the generation of the relationship between all existing physical input parameters found in the data cluster and the power output (*P*) parameter. In addition, in order for the computer to make the best calculations and simulate as quick as possible; considering the physical units of input and output parameters, the most appropriate physical units were chosen. In this regard, the unit for the temperature (*T*) parameter of the wastewater was selected as Kelvin (K), the unit for its conductivity was chosen as µS/µm, the unit for the degree of acidity was considered as dimensionless, the unit

for the daily cumulative flow of the methane gas to the gas generator was taken into account as m^3/day, and the power (P) obtained from the gas generator was distinguished as MWh/day, and the simulation calculations were carried out accordingly. Another important factor that affects the sensitivity of the study and increases the prediction success by reducing error values or bringing the correlation coefficient (R) value closer to 1 is the percentage and number of the test data used in the training of the algorithm. Generally, 80% of the total data in the dataset is used for training, while the remaining 20% of the cumulative data is used for testing. Although the percentages are adjusted in this way, the large number of the total data considered in estimations is important for the creation, development and the improvement of the mathematical algorithm. In case of insufficient data, mathematical modeling is usually not sufficient for simulation and unqualified predictions with high error values or low correlation coefficients (R) may occur. For this purpose, the number of the samples of actual observed experimental training and testing data should be increased as much as possible in order to further reduce the error rates or to increase the correlation coefficient values. In this way, different physical input parameters in the data cloud are learned and associated better by the artificial intelligence algorithm, with respect to the output parameter; and as a result, highly successful prediction results that are closely approaching to the real values could be obtained. In the current study, different types of modelling combinations were created, a variety of different numbers of hidden layers (HL) were tested, as well as best physical units of the physical parameters were adjusted. The remaining regulation would be only on the increase of the total number of the data samples, considered in the computations, in order to obtain more accurate predictions.

It may be necessary to deal with complex relationships and differential equations in analytically relating an important physical quantity such as power output (P) to other known physical parameters. In some cases, it may not be possible to solve these equations using analytical methods. Whereas, in many of the cases, complex, analytical and non-stochastic problems cannot be solved directly. Besides, to solve these numerically, advanced computers are needed and it is usually quite time consuming. On the other hand, in some cases, a direct relationship cannot even be established between input and output parameters. In determining the relationship between them, artificial neural networks (ANNs) provide fast, precise and highly sensitive results. In the ANNs, generally, the output parameter is easily obtained by a mathematical relationship in which the input parameters are substituted to their proper places in an exponential function. The algorithm is easier to design and use than the other methods. For example, in this study, technical information of the facility is not needed in estimating the power output (P) obtained by combusting the methane gas in the gas generators, in which the methane gas was acquired due to the decay of the wastes of the waste-water in the considered wastewater treatment facility. On the other hand, in complex analytical relations or empirical equations, either further technical knowledge of the facility is required or more physical input parameters and measurements on those are needed. Whereas here, in the current study, with artificial intelligence (AI), the output power (P) parameter is only and rapidly associated with the available physical and chemical parameters found in the data cluster including temperature (T), degree of acidity (pH), conductivity (σ), and the methane gas daily consumption (Q).

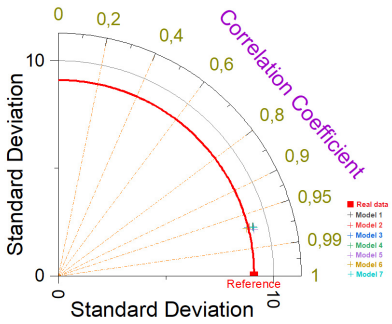

Figure 18. Taylor diagram for the output power predictions (P)

The correlation coefficient (R) values and the standard deviations of the best sub-models, in which the best sub-models of each main models were presented in Tables 1 and 2 using the bold

color, are demonstrated in Taylor diagram, which is shown in Figure 18. In this regards, the results of 5 HL of Main Model 1, 9 HL of Main Model 2, 6 HL of Main Model 3, 10 HL of Main Model 4, 9 HL of Main Model 5, 8 HL of Main Model 6 and 5 HL of Main Model 7 have been indicated in this figure. It is concluded that the results of the standard deviations and the correlation coefficients (R) are more or less close to each other for these best sub-models. Accordingly, the correlation coefficient range for these best sub-models were reported to be in the range of $0.9691 \leq R \leq 0.9730$, as shown in Tables 1 and 2 as well as on Figure 18. On the other hand, for the cited best sub-models of main Models 1, 2, 3, 4, 5, 6, and 7; the standard deviation values were reported to correspond respectively to 9.3186 MWh/day, 9.1563 MWh/day, 9.0878 MWh/day, 9.2934 MWh/day, 9.3248 MWh/day, 9.0284 MWh/day, and 9.3278 MWh/day. Besides, the standard deviation of the real measured output power (P) was concluded to be 9.0987 MWh/day. The standard deviation values of all these best prediction sub-models being close to each other as well as being close to the standard deviation value of the real observed output power (P) data represent the high accuracy of the predictions. Besides, high correlation coefficients (R) dispersed so close to each other on the Taylor diagram show the high compatibility of the power data estimations with respect to the real observed power (P) data.

4. CONCLUSION

In this study, data predictions on the daily total energy generation (P) of a gas generator of an installed waste water treatment plant, using the physical and chemical parameters of the waste water of this installed waste water treatment plant were performed; using the techniques of artificial intelligence methods.

In this context, a cumulative of 7 main models and 56 sub-models were created. The ANN predictions indicated that best prediction outcome was obtained with the utilization of 9 *HL* of main Model 5. This best model was obtained by adjusting three parameters as input, in which three of them were selected as physical parameters, and no chemical parameter was utilized. Among those physical parameters, the temperature (T), the electrical conductivity (σ), and the daily consumption of the methane gas (Q) of the facility are taken into account. Besides, this superior model of ANN had given statistical error outcomes of 6.1279% MAPE, 2.1540 MWh/day RMSE, and 0.9730 R, for the prediction of the physical parameter of the gas generator daily total energy generation (P).

Accordingly, when three statistical error parameters are taken into account, it was observed and concluded that at this best model of artificial intelligence tool of ANN, close and accurate results of predictions could be acquired for power generations (P) of the gas generator of the facility.

5. ACKNOWLEDGEMENTS

Thanks to the facility of the waste-water treatment plant and the staff providing this data, in conducting of this study.

6. REFERENCES

1. Ilhan, A., 2022. Forecasting of River Water Flow Rate with Machine Learning. Neural Computing and Applications, 34, 20341-20363.
2. IEA, 2023. International Energy Agency 2023, https://www.iea.org/.
3. Zhang, J., Yan, J., Infield, D., Liu, Y., Lien, F., 2019. Short-term Forecasting and Uncertainty Analysis of Wind Turbine Power Based on Long Short-term Memory Network and Gaussian Mixture Model. Applied Energy, 241, 229-244.
4. Jung, J., Broadwater, R.P., 2014. Current Status and Future Advances for Wind Speed and Power Forecasting. Renewable and Sustainable Energy Reviews, 31, 762-777.
5. Tascikaraoglu, A., Uzunoglu, M., 2014. A Review of Combined Approaches for Prediction of Short-term Wind Speed and Power. Renewable and Sustainable Energy Reviews, 34, 243-254.
6. Liu, H., Tian, H., Li, Y., 2015. An EMD-recursive ARIMA Method to Predict Wind Speed for Railway Strong Wind Warning

System, Journal of Wind Engineering and Industrial Aerodynamics, 141, 27-38.

7. Kavasseri, R.G., Seetharaman, K., 2009. Day-Ahead Wind Speed Forecasting Using F-ARIMA Models. Renewable Energy, 34(5), 1388-1393.

8. Khosravi, A., Koury, R.N.N., Machado, L., Pabon, J.J.G., 2018. Prediction of Hourly Solar Radiation in Abu Musa Island Using Machine Learning Algorithms. Journal of Cleaner Production, 176, 63-75.

9. Shi, X., Lei, X., Huang, Q., Huang, S., Ren, K., Hu, Y., 2018. Hourly Day-ahead Wind Power Prediction Using the Hybrid Model of Variational Model Decomposition and Long Short-term Memory. Energies, 11(11).

10. Zhang, Z., Ye, L., Qin, H., Liu, Y., Wang, C., Yu, X., Yin, X., Li, J., 2019. Wind Speed Prediction Method Using Shared Weight Long Short-term Memory Network and Gaussian Process Regression. Applied Energy, 247, 270-284.

11. Ilhan, A., Bilgili, M., Sahin, B., Akilli, H., 2015. Estimation of Aerodynamic Characteristics for a Horizontal Axis Wind Turbine. International Journal of Engineering and Natural Sciences, 9(2), 51-57.

12. Bilgili, M., 2010. Prediction of Soil Temperature Using Regression and Artificial Neural Network Models. Meteorology and Atmospheric Physics, 110(1), 59-70.

13. Graupe, D., 2007. Principles of Artificial Neural Networks. World Scientific Publishing Co. Pte. Ltd., 2nd ed., USA.

14. Bilgili, M., Sahin, B., 2010. Comparative Analysis of Regression and Artificial Neural Network Models for Wind Speed Prediction. Meteorology and Atmospheric Physics, 109(1), 61-72.

15. Tchobanoglous, G., Burton, F.L., Stensel, H.D., 2003. Wastewater Engineering: Treatment and Reuse. Metcalf & Eddy, Inc., McGraw-Hill, 4th ed., New York, USA.

16. Ilhan, A., 2023. Energy Recovery Analysis of a Biomass Type of Installed Waste Water Treatment Plant. Cukurova University Journal of the Faculty of Engineering, 38(1), 169-183.