ZANCO Journal of Pure and Applied Sciences

TYPE Original article
DOI http://dx.doi.org/10.21271/ZJPAS.37.5.9

A Machine Learning Models for Predicting Power Output in Gas Turbines: A Case Study from Perdawood Power Station, Erbil, Iraqi Kurdistan

OPEN ACCESS

*Corresponding author

Asaad Saber Hamad Ameen
emgasaad86@gmail.com

RECEIVED :27/12/2024
ACCEPTED :14/05/2025
PUBLISHED :31/10/2025

KEYWORDS:
Gas Turbine (GT), Machine Learning (ML), Support Vector Regression (SVR), Power Generation Forecasting, Dataset.

Asaad Saber Hamad Ameen*, Ibrahim Ismail Hamarash

Department of Electrical Engineering, College of Engineering, Salahaddin University-Erbil, Erbil ,Kurdistan Region, Iraq

ABSTRACT

Gas Turbines (GT) are essential generation units in both centralized and distributed power systems. In many regions worldwide, GTs are the primary source of electrical energy due to their affordability, efficiency, and flexibility. GT can quickly reach peak output, making them ideal as auxiliary units, additionally, they are often integrated with renewable energy sources to improve system adequacy and reliability. Accurate forecasting of GT power output is crucial for achieving efficient management and smart operation of these units.

This study presents the first application of machine learning techniques for electrical power output prediction from a combined cycle gas power plant in the Iraqi Kurdistan region. A novel dataset comprising 7,893 data points collected during base load operations over four years, from 2019 to 2023 at the Perdawood Electrical Power Station in Erbil, Kurdistan Region of Iraq. The dataset includes eight features - Ambient Temperature (AT), Ambient barometric Pressure (AP), Specific Humidity (CH), grid frequency (DF), Gas Fuel Flow (GP), Compressor Pressure Discharge (CPD), Compressor Pressure Ratio (CPR), and Exhaust Temperature (TTXM), and the Electrical Power Output (EP) as the target variable.

The dataset has been preprocessed, and the Support Vector Regression (SVR) machine learning algorithm is applied to models and trains the dataset. The model's accuracy is carefully evaluated using the effective metrics such as Mean Squared Error (MSE), R-squared (R^2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The results demonstrate the effectiveness of the Support Vector Regression (SVR) algorithm for forecasting the power output of gas turbines, that authenticates the practicability and value of applying machine learning to real-world GT operation in the Kurdistan region. This work provides a strong foundation for introducing a new strategy of smart energy management in the power sector in the region.

1.Introduction

Gas turbines (GT) are a major source of electrical energy production in many regions worldwide. Particularly in nations switching from conventional fossil fuel systems to more environmentally friendly and economically viable energy sources, such as a GT with a better thermal efficiency and lower CO2 emissions than other fossil fuel power plants. Modern advancements in machine learning, real-time monitoring, and CCGT technology have thrown up new opportunities to improve power plant performance (Liu and Karimi, 2020, I Blinov and Deryabin, 2022).. The energy industry is under pressure to increase efficiency, include renewable sources, and satisfy the rising demand for power, therefore the improvement is pertinent in places like Kurdistan, Iraq. GT uses gas fuel as the fuel for combustion. The turbine contains an air inlet system called a filter house, an axial air compressor, combustion section, turbine, exhaust, and generator. One of the gas turbine foremost components is the air inlet filter system, which is used to supply the filtered and cleaned airflow to the compressor and turbine, it's used to decrease the compressor air inlet acoustical level(Mohapatra and Sanjay, 2014).

Accurate power-generating prediction is vital for the power systems' efficient management and operation. It's important for power system grid stability that helps to sustain the stability between power generation and demand, cost-effective and economic operation and efficiency, plan for operative integration with renewable energy sources such as wind plant or solar plant, and a complete operational arrangement for maintenance activities. Furthermore, forecasting power generation helps the utilities in their operation plan to meet the monitoring emission goal and work to use cleaner energy optimization(Abad et al., 2020).

Artificial Intelligence (AI) and its branch, Machine Learning (ML), are currently the most widely used techniques for predictions in the electrical energy industry. ML forecasting algorithms have proved significant potential in enhancing the sustainability, reliability, and efficiency of power systems (Kabengele et al., 2023, Kaledio and Sustainability, 2024). The fundamental concept behind forecasting is identifying patterns within historical data that influence power output. These patterns and their values can then be used to predict future power output levels. These patterns are called features in ML taxonomy and their selection is carried out with the help of domain experts. The historical data or the dataset consists of many observations of these features. This dataset then is fed to ML algorithms for building a prediction model. Many well-established algorithms exist in the literature of ML and power forecasting. Machine learning algorithms are commonly employed for datasets with a known target, also known as supervised continuous-data problems(Farooqi et al., 2019, Rawal and Ahmad, 2022). The (SVR) model and other ML techniques applied for electric power demand forecasting, the result shows that the (SVR) model is highly precise in estimating the electric power demand. The result for errors such as MSE and MAE is achieved best by applying the SVR Model (Chen et al., 2023, Hasan et al., 2025).

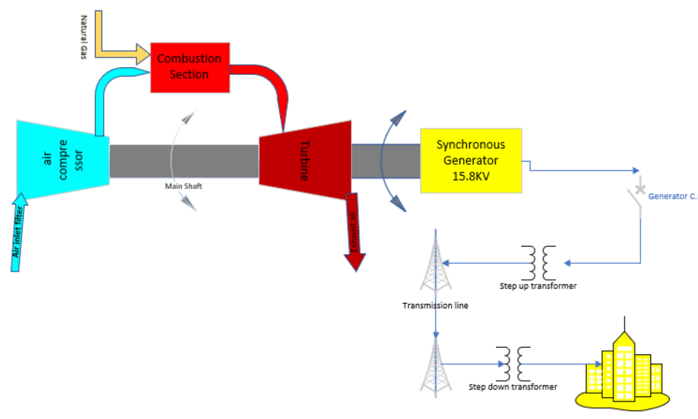

Figure 1. Gas turbine power generation process

The ML Regression algorithm has been used to forecast the electrical power output of CCGT, resulting in improved system efficiency, increased MW output, and more economical operation of the system (Lobo et al., 2019). Using ML in GT power prediction is questionable because of the nonlinearity of the features that contribute to output power levels. Domain experts suggested using features such as ambient air temperature, barometric pressure, related humidity, and vacuum to expect the combined cycle power output. These features have been modeled and evaluated with different successful rates according to the dataset accuracy and integrity in several studies (Alketbi et al., 2020, Santarisi and Faouri, 2021). In (Shaik et al., 2024) The robust prediction model is developed and evaluated by using the Extreme Gradient Boosting (XGBoost) algorithm. The model is built to forecast the gas turbine power output by using the high-resolution historical dataset comprising together the environmental and operational parameters. The main objective is to improve energy generation efficiency and maintenance strategies plan. model performance was evaluated via using effective matrices like R-squared (R^2) and RMSE, MSE, and MAE. The results illustrate the model's accurateness and dependability, designing the solid foundation for real-time integration of the implemented XGBoost algorithm and the GT control system.

The hydropower plant electrical power output forecasting is proposed in (Qi et al., 2025) by using difference factors such as weather, water storage and the power production, The study is implemented with SVR and XGBoost algorithms separately and with hybrid models. Performance of the model is assessed and optimized with R-squared (R^2) and RMSE, The XGBoost indicated the higher performance, and it's revealed the most accuracy. the prediction model for carbon monoxide (CO) and nitrogen oxide (NOx) output from combined cycle gas turbine is presented in (Pachauri, 2024). The work objective is to create a stacked ensemble machine learning (SEM), the performance is compared with SVR , decision tree and linear regression. The result shows that SEM relate to other ML method can reduce the errors RMSE 5.7–93.8% for NOx and 1%-41.5% for CO.

For Photovoltaic (PV) power forecasting the regression ML (SVR) is used in (Das et al., 2017), the power output from PV, atmospheric data, and weather condition data were collected and used for predicting the result. It shows that the applied model is best achieved and has lower numerical complexity.

An effective and reliable prediction and estimation model for a combined cycle power plant's (CCPP) full load power generation is proposed using the ML transformer encoder and deep neural network (DNN) (Yi et al., 2023). This paper introduces a new model for predicting power output in the power generation division. The prediction of gas turbine performance and characteristics using two machine learning models is presented in (Liu and Karimi, 2020), In the study by (Wankhede and Ghate, 2018) regression models and Artificial Neural Networks (ANN) were enhanced to forecast air compressor and turbine performance. The results from using both techniques showed an error of less than 1%. In (González-Briones et al., 2019) Electrical consumption prediction by using the ML models is proposed, in the review results show that using the Linear Regression and (SVR) model for predicting the dataset, gets the higher accuracy and best results.

For renewable power generation integration, the ANN, (SVR) and Gaussian Process Regression (GPR) are applied (Sharifzadeh et al., 2019), and all three models were found proficient in forecasting wind and solar power generation. Also, found that the forecast progression for electrical demand, wind, and solar energy power generation was dissolute precisely adequate to successfully change the alternative electricity sources.

This study novelty is to build and evaluate ML model using an extensive, high-frequency operational dataset uniquely sourced from a real-world combined cycle gas power plant in Iraq. The dataset is preprocessed and standardized for ML applications. The results of this study will help in developing and operating a smart management system for the station, facilitating pre-planning to balance demand and generation throughout the year. Additionally, it will help reduce operational costs, thereby affecting the overall cost of electricity tariffs nationwide. The established dataset includes eight features, the ambient temperature, pressure, specific humidity, gas fuel flow, compressor inlet pressure, compressor outlet pressure, and compressor pressure ratio. Data has been collected over four years at three-hour intervals using multiple sensors around the gas turbine power plant.

The structure of the paper is systematized as follows: Section one introduces the concept of gas turbines (GT), the selection of features, and machine learning (ML), along with a review of related works. Section two outlines the methodology for feature selection and data collection. Section three presents the SVR algorithm, and the implementation of ML. Section four is dedicated to the evaluation and discussion of the results, and the final section provides the conclusion.

2. Materials and methods

This section details the methodology followed in this study, encompassing standard machine learning steps: feature selection, data collection and preprocessing, model training and finally evaluation of the model performance.

2.1 Feature Selection

The Gas Turbine Generator (GTG) is an electrical power generation system that uses a gas turbine to produce electricity. It comprises three main sections; the compressor section, the turbine and combustion system, and the auxiliary systems, The compressor section contains the multi-stage axial compressor, while the turbine and combustion system include the three-stage turbine and 12 combustion chambers where air, fuel, and a spark mix to produce combustion, and the auxiliary systems include components such as the lubrication oil system, cooling water system, and fuel system. The generator system consists of the synchronous generator and the excitation system. The GTG also features a control system that manages the entire gas turbine operation, including startup, shutdown, synchronization with the electrical grid, and monitoring and controlling various instruments such as motors, current transformers (CT), voltage transformers (VT), and pressure, level, and flow measurements. The GTG component diagram is shown in Figure 1.

The gas turbine operates on the Brayton cycle, also known as the Joule cycle. In this cycle, air is pressurized by the compressor, and fuel is also pressurized. In the combustion section, the air and fuel are burned, passing through the gas turbine and then exiting through the exhaust. The Brayton cycle process illustrates the transmission of electricity from the generator to homes via circuit breakers and transformers (Boyce, 2011, Ibrahim and Rahman, 2012).

Table 1 The datasheet of a gas turbine in Perdawood CCGS, Erbil

Manufacture and Type	GE 9171E
Generator	Synchronous Generator
Exciter	Brushless exciter
Rated capability	125 MW (designed output)
Air Inlet Conditions	Compressor inlet pressure 66 mmwc, temp. 7c
Compressor Stages	17
Compressor Pressure Ratio	12.3/1
Combustor	DLN 1.2
Fuel	Natural Gas
Power Turbine stages	3
Exhaust Gas Flow	1410.8t/h
Exhaust Gas Temp	553.8℃
Starting	Electric Motor
Control system	SPEEDTRONIC MARKVI GTG control system

Feature selection was guided by analyzing several operational factors and a comprehensive understanding of gas turbine operational behavior, both environmental and operational variable known to have a direct influence on power output ere prioritized to ensure that the model captures key dynamics affecting turbine performance: (Niazy et al.). The main factors with electrical power output from GT depend on:

1. Weather Condition: Weather conditions refer to the external environmental factors that can

influence the performance and efficiency of the turbine. Key weather conditions include:

Ambient Temperature (AT): The temperature of the air surrounding the gas turbine. Higher ambient temperatures can reduce the air density, leading to a decrease in the mass flow rate through the turbine, which can lower the power output.

Ambient Barometric Pressure (AP): The atmospheric pressure of the environment where the gas turbine is operating. Changes in barometric pressure can affect the air intake and combustion process, impacting the turbine's performance.

Specific Humidity (CP): The quantity of water vapor is present in the air. High humidity levels can affect the combustion process and the efficiency of the turbine, as moist air has a different heat capacity compared to dry air.

2. Fuel gas flow: the flow rate of the natural gas provided to the gas turbine for ignition. Appropriately controlling and monitoring the fuel gas flow and quality are critical for sustaining effective economic operations.
3. Compressor discharge pressure: the output pressure of the compressor is one of the factors that affect proportionally the gas turbine generator's electrical output.
4. Grid frequency: it’s the electricity frequency of the entire country. When the power generation is more than the power demand the grid frequency is increased. Vice versa, if the demand is more than the generation the grid frequency falls.
5. Compressor air ratio: It shows the efficiency of the axial air compressor which is the main part of the gas turbine generator. CPR is a generated parameter and is affected by CPD, compressor inlet pressure, and barometric pressure.
6. Turbine output temperature (exhaust temperature): Exhaust temperature (TTXM): Its turbine outlet temperature, the gas turbine is the temperature control turbine. That means in the summer and winter the temperature of the exhaust is close to fix. Other parameters will change based on the exhaust temperature. The power output will be one factor that is affected by the exhaust temperature. Table 1 shows a parameter for a Gas Turbine for Perdawood CCGS in Erbil.

The weather conditions directly influence the air intake, combustion efficiency, and overall power output of the system. Managing the fuel gas flow and maintaining the proper specifications contribute to an efficient and stable operation of gas turbines, ensuring high power output, and minimizing emissions. The compressor pressure ratio plays a significant role in the efficient and reliable operation of gas turbines. The inlet air temperature is another important variable for maximizing the efficiency and power output of the turbine. These seven factors have been selected as features for the dataset, and the power output is chosen as the target variable, the correlation between features and target is shown in heatmap figure 2.

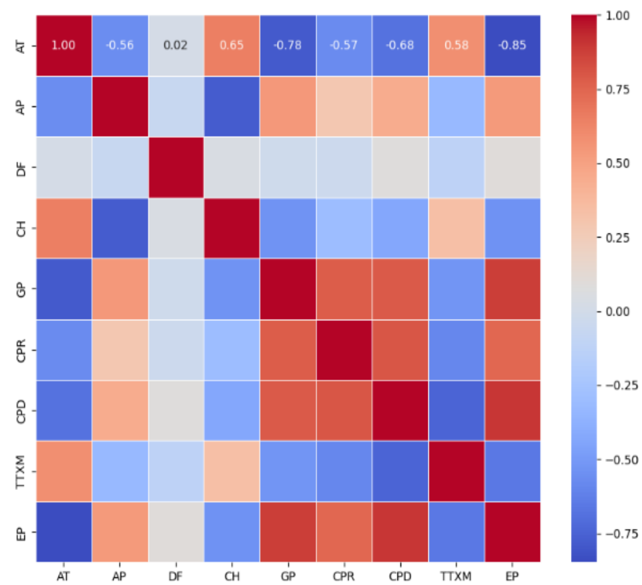

Figure 2. correlation heatmap

2.2 Data Collection and Preprocessing

To provide comprehensive and precise measurements, various kinds of sensors and devices are used during the process of collecting data. Below are the types of sensors used for:

- For ambient temperature, the Resistance Temperature Detector (RTD) type PT100 is used.

- For turbine outlet (exhaust) temperature, the thermocouple type K is used.
- Specific Humidity is the output calculation for ambient temperature and relative humidity measurement.
- Ambient barometric Pressure and Compressor Outlet (discharge) Pressure is measured by the Rosemont pressure transmitter.
- Natural gas flow is measured by the Rosemount Coriolis Flow meter.
- Electrical power output and grid frequency are measured by the Generator protection system GE G60.

The dataset contains eight features which are ambient temperature (AT), ambient barometric pressure (AP), grid frequency (DF) specific humidity (CH), gas fuel flow (GP), compressor discharge pressor (CPD), compressor pressor ratio (CPR), exhaust temperature (TTXM) and one target that is the electrical power output (EP).

The data was collected over a span of four years, with samples taken at 40ms, but because of the dataset size, it's chosen for three-hours interval during preprocessing. Data collection occurred throughout all four seasons to observe how weather conditions and other factors influenced the parameters.
In the initial data cleaning process, the absence of missing values where removed. Furthermore, abnormalities due to the electrical grid instability or load transitional process where the dataset were manually filtered to ensure consistency of base-load operation data. The following preprocessing steps were applied using python pandas and sci-kit-learn libraries:

- Outlier removal: outlier due to grid instabilities, load transitional process, maintenance, or stop operation status were removed.
- Feature scaling: all the eight features and target were normalized using Min-Max scaling to transform data into the [0,1] range, to ensure its matching SCR input.
- Feature engineering: no synthetic features were introduced in the final model.
- Data splitting: the dataset was split into testing (20%) and training (80%) sets using graded random sampling to preserve data distribution. It ensures that both subsets precisely represent the operational condition of the four-year data for gas turbines.
- Cross-validation: A 5-fold cross-validation strategy was applied on the training set during parameter tuning. to prevent overfitting and ensure generalizability.

Figure 3 shows the distribution and range for each feature in the boxplot, the median, interquartile range, minimum, and maximum are illustrated for each feature, and Figure 4. Shows the header for the dataset.

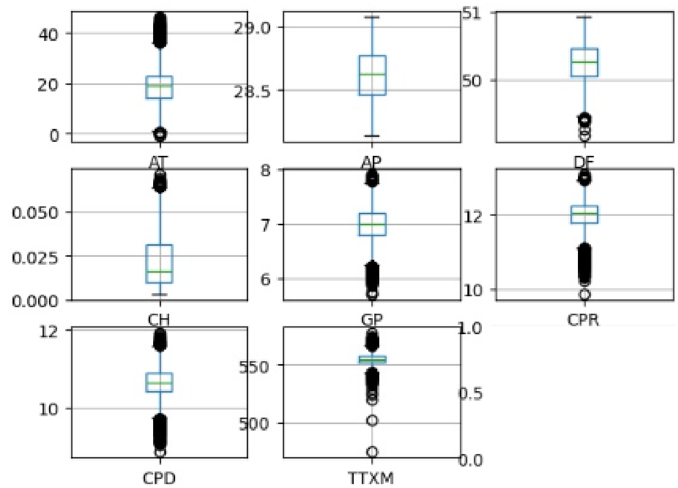

Figure 3. The vertical boxplots for all features

```
In [169… dataset = pd.read_csv("D:/powerone.csv")
In [170… dataset
Out[170…
```

	AT	AP	DF	CH	GP	CPR	CPD	TTXM	EP
0	15.31	979.41	50.08	0.01	7.28	12.80	11.31	548.68	119.15
1	17.75	979.03	49.67	0.01	7.26	12.62	11.10	551.71	117.35
2	14.51	978.34	50.65	0.01	7.51	12.67	11.44	539.46	120.32
3	10.48	977.65	49.77	0.01	7.58	12.83	11.37	547.85	121.71
4	8.93	976.96	50.30	0.01	7.64	12.98	11.61	543.26	125.33
...	...	...	...	...	...	...	...	...	...
7874	15.10	969.10	50.81	0.02	7.20	11.91	10.60	554.23	117.17
7875	18.02	968.96	50.03	0.02	6.70	11.93	10.34	559.39	111.87
7876	20.79	968.83	49.81	0.01	6.77	12.01	10.22	561.39	110.86
7877	18.06	968.44	50.19	0.02	6.94	12.09	10.45	557.77	113.55
7878	18.65	967.98	50.24	0.03	7.09	12.07	10.42	558.04	113.24

7879 rows × 9 columns

Figure 4. The header for the dataset

2.3 Support Vector Regression Algorithm

The established dataset contains eight features, all of them with numerical data. The target is output power, which consists of continuous numerical data. This kind of problem is classified as a Continuous Supervised Machine Learning problem (CSML). There are many ML algorithms in the literature to solve CSML problems. An insight into the behaviors of the data reveals that higher temperatures reduce air density, affecting combustion efficiency and power output. The relationship between ambient temperature and power output is non-linear. The impact of humidity on power output is also non-linear. High humidity levels can affect the combustion process, but the relationship depends on various factors such as the design of the turbine and the combustion system. On the side, changes in pressure affect air density and the mass flow rate through the turbine, influencing power output. The relationship between compressor pressure ratio and power output can be non-linear, especially at higher pressures where the system may reach operational limits. While the basic relationship between fuel flow rate and power output might appear linear (more fuel generally produces more power), the actual relationship can be non-linear due to factors such as combustion efficiency, heat losses, and operational constraints of the turbine(Solyali, 2020, Abad et al., 2020).

The combined effects of these parameters typically result in non-linear relationships with power output. Due to these non-linear relationships, Support Vector Regression (SVR) is the most proper algorithm for nonlinear continuous supervised problems like our problem.

The (SVR) in sci-kit-learn is a part of the 'sklearn', it's responsible for running the support vector machine (SVM) algorithms for regression, outlier detection, and classification of supervised machine learning(Mardin Abdullah et al., 2024). The Support Vector Regression technique is implemented by the SVR module, which also handles a variety of kernel functions, such as radial basis function (RBF), polynomial, and linear kernels(Hong, 2009).

The (SVR) algorithm was found by Vapnik. SVR aims to discover a function $f(x)$ that is as flat as feasible and has a maximum ϵ (epsilon) variation from the actual target values for all training data(Rawal and Ahmad, 2022). The f(x) is clarified in equation 1.

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_i, x) + b \dots\dots\dots\dots\dots 1$$

Where α_i and α_i^* Are the Lagrangian multipliers $K(x_i, x)$ Is the kernel function, b is the bias term and the support vectors are expressed by x_i. The SVR model's overall error minimization mathematical equation is illustrated in equation 2, and there are some constraints and limitations for the data tuning:

$$\frac{1}{2} \| \mathbf{w} \|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \dots\dots\dots\dots\dots\dots\dots 2$$

Limitations are: $y_i - (\mathbf{w} \cdot \phi(x_i) + b) \leq \epsilon + \xi_i$

$$(\mathbf{w} \cdot \phi(x_i) + b) - y_i \leq \epsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0$$

Where, $\phi(x)$ is the features mapping to a higher-dimensional space, and C is the regularization parameter. Hence, to use the function to make predictions for a new input x the support vector regression (SVR) model is mathematically shown in equation 2 in the dual space.

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \exp\left(-\gamma \|x_i - x\|^2\right) + b \dots\dots\dots 3$$

Where γ in the kernel function is a parameter that defines the width of the Gaussian function.

Different types of kernel functions exists. The kernel function simply uses the input data and transforms it into the favorite form. In this paper, due to easier implementation and capability to map the nonlinearity of training data into the higher dimensional space the Gaussian radial basis function (RBF) is used. Furthermore, the Gaussian RBF is appropriate to work with nonlinear relationships problems(Hong, 2011, Chiman Haydar and Abbas, 2022). The flowchart of the SVR algorithm is illustrated in figure 5

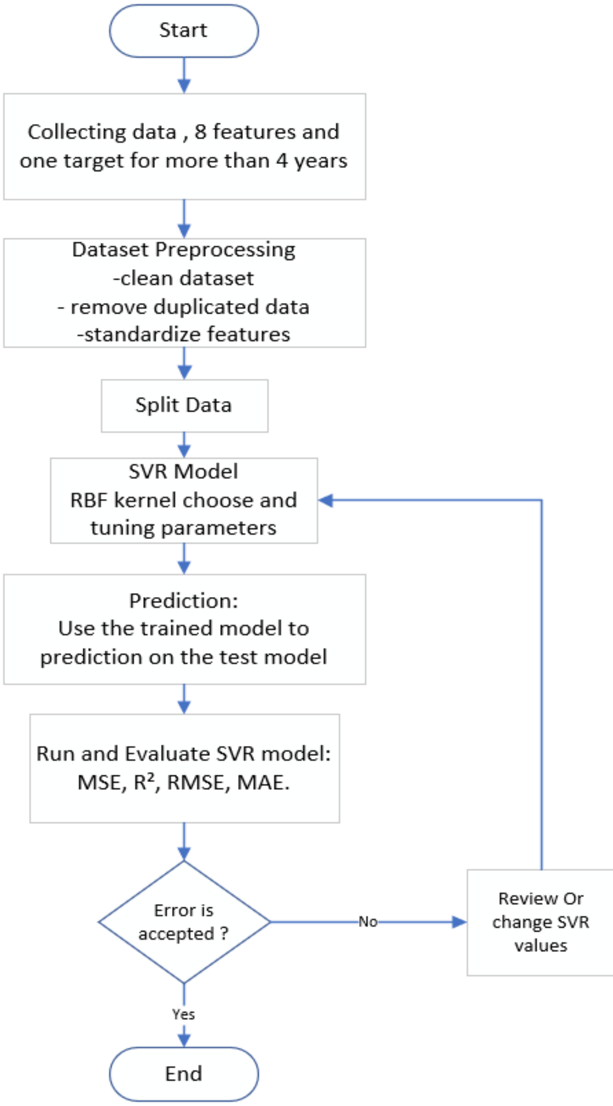

Figure. 5 SVR algorithm

2.4 Evaluation

An SVR (Support Vector Regression) model's performance on the test data is evaluated using a variety of metrics. Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R2 score) are typical metrics used in evaluating regression models. Scatter plots of actual vs. projected values are one type of visualization that can shed light on how well the model works(Haider Abdula, 2023).

Means square error: It is the squared average of the predicted and actual values:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \breve{y}_i\right)^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (4)$$

n is the number of data

y_i actual value

$\breve{y}_i$ predicted value.

Root Means Squared Error: is the square root of MSE and shows the error average value.

$$RMSE = \sqrt{MSE} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. \quad (5)$$

R-squared or R²: is from 0 to 1, based on the value from 0 to 1 shows how the dependent variable change is shown by an independent variable, or it indicates how good fit of the regression model to the dataset. 1 shows perfect fit and 0 shows no fitted target with the features.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \breve{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

Where $\bar{y}$ is the actual value

Means Absolute Error (MAE): it's a metric of the error average value without a square on the actual and prediction value.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \breve{y}_i\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (5)$$

3. Results and Discussion

In this study, we used the SVR class from the sklearn Python library to perform Support Vector Regression (SVR) for predicting the power output of a Gas Turbine (GT). To ensure the robustness of our model, the dataset was divided into training (80%) and testing (20%) sets used to avoid overlapping in time series datasets. The training data was employed to build the SVR model, although the testing part was used to evaluate its performance. Key parameters of the SVR class such as the kernel type (RFB), regularization parameter (C=100), and gamma (gamma=0.1) were carefully tuned by using the GridSearchCV from scikit-learn library to optimize the model's performance and accuracy, that shown an in-depth search over a detailed parameter grid with 5-fold cross validation the scatter plot comparing actual vs. predicted values is shown in figure 6. .

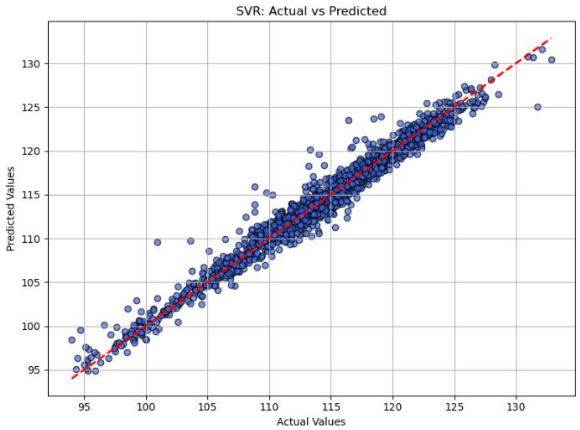

Figure 6. scatter plot for actual vs. predicted values

The SVR model's performance was evaluated by using several metrics:

R-squared (R^2): 0.96638 including R-squared (0.96638). It's shown that the model justifies almost 96.6% of the variance in power output, which proves an excellent fit of the data.

Mean Squared Error (MSE): 1.41. It measures the mean of the squares of the errors. A low value indicates the model's predictions are close to the actual values.

Root Mean Squared Error (RMSE): 1.187. As the square root of MSE, this metric maintains the same unit as the target variable and confirms that the prediction errors are small.

Mean Absolute Error (MAE): 0.847. Measurement is very small, signifying that the model is accurate and precise.

The SVR model displayed excellent predictive accuracy across all evaluation metrics, indicating its reliability for predicting gas turbine power output using the chosen environmental and operational variables. These findings highlight the potential of machine learning to improve forecasting precision and support more efficient operational planning in CCGT power plants.

4. Conclusion

Gas turbine generators are among the most often used kind of power plant in Kurdistan Region of Iraq. About 70% of the local power plants use this technology. Improving both efficiency and electrical generation capacity depends on the elements affecting gas turbine power production. To connect important environmental and operational variables with the power output of gas turbine generators, this paper presented a novel, high-resolution dataset gathered over four years. This dataset, to our knowledge, is the first of its like in the area and offers insightful analysis for academics and business experts. The power output was predicted by Support Vector Regression (SVR) using this dataset. The model showed good predictive accuracy with low MAE, MSE, and RMSE values and a high R^2 score, hence validating its appropriateness for practical use.

The findings showed that power production is greatly influenced by meteorological conditions, especially ambient temperature, barometric pressure, particular humidity, and compressor pressure ratio. Identified as successful ways to increase generating efficiency were operational tactics included the use of evaporative coolers and wet compression devices to regulate ambient temperature. Furthermore, combining machine learning-based predictive models can help to optimize operational costs, maintenance scheduling, and real-time load balancing in combined cycle gas turbine (CCGT) facilities.

Relying on data from one plant and one kind of gas turbine generator, ignoring dynamic operational variables like turbine age, maintenance history, and component lives, may compromise the validity and prediction accuracy of the model under more complicated situations.

Acknowledgement

The authors would like to extend their heartfelt thanks to the Erbil Perdawood Power Plant for their unwavering support during the data collection phase of this study. We greatly appreciate the cooperation and assistance provided by the plant's management and staff, who ensured that all necessary resources and information were made available.

The authors extend their appreciation to the Department of Electrical Engineering, College of Engineering at Salahaddin University-Erbil, for providing access to laboratories and facilities that supported the experimental work in this study.

Conflict of interest

"The authors declare that there is no conflict of interest regarding the publication of this paper. No financial or personal relationships have influenced the research, analysis, or interpretation of the results presented in this study."

References

Abad, L. A., Sarabia, S. M., Yuzon, J. M. & Pacis, M. C. A Short-Term Load Forecasting Algorithm Using Support Vector Regression & Artificial Neural Network Method (SVR-ANN). 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC), 8-8 Aug. 2020 2020. 138-143.

Alketbi, S., Nassif, A. B., Eddin, M. A., Shahin, I. & Elnagar, A. Predicting the power of a combined cycle power plant using machine learning methods. 2020 International conference on communications, computing, cybersecurity, and informatics (CCCI), 2020. IEEE, 1-5.

Boyce, M. P. 2011. Gas turbine engineering handbook, Elsevier.

Chen, G., Hu, Q., Wang, J., Wang, X. & Zhu, Y. 2023. Machine-Learning-Based Electric Power Forecasting. 15, 11299.

Chiman Haydar, S. & Abbas, M. A. 2022. Comprehensive Study for Breast Cancer Using Deep Learning and Traditional Machine Learning. Zanco Journal of Pure and Applied Sciences, 34, 22-36.

Das, U. K., Tey, K. S., Seyedmahmoudian, M., Idna Idris, M. Y., Mekhilef, S., Horan, B. & Stojcevski, A. 2017. SVR-Based Model to Forecast PV Power Generation under Different Weather Conditions. 10, 876.

Farooqi, B. A., Kazmi, A. A. & Janjua, A. K. Machine learning in power markets. 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2019. IEEE, 1-6.

González-Briones, A., Hernández, G., Corchado, J. M., Omatu, S. & Mohamad, M. S. Machine Learning Models for Electricity Consumption Forecasting: A Review. 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), 1-3 May 2019 2019. 1-6.

Haider Abdula, H. 2023. Online Student Feedback System Using Machine Learning. Zanco Journal of Pure and Applied Sciences, 35, 78-85.

Hasan, M. S., Tarequzzaman, M., Moznuzzaman, M. & Ahad Juel, M. A. 2025. Prediction of energy consumption in four sectors using support vector regression optimized with genetic algorithm. Heliyon, 11.

Hong, W.-C. 2009. Hybrid evolutionary algorithms in a SVR-based electric load forecasting model. International Journal of Electrical Power & Energy Systems, 31, 409-417.

Hong, W.-C. J. E. 2011. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. 36, 5568-5578.

Ibrahim, T. K. & Rahman, M. J. I. J. O. E. E. 2012. Effect of compression ratio on performance of combined cycle gas turbine. 2, 9-14.

Kabengele, K. T., Olayode, I. O. & Tartibu, L. K. J. a. S. 2023. Analysis of the Performance of a Hybrid Thermal Power Plant Using Adaptive Neuro-Fuzzy Inference System (ANFIS)-Based Approaches. 13, 11874.

Kaledio, P. J. R. & Sustainability. 2024. Machine Learning Applications in Electric Power Systems: Enhancing Efficiency, Reliability, and Sustainability.

L Blinov, V. & Deryabin, G. Estimation of gas turbine technical condition using machine learning methods. IOP Conference Series: Earth and Environmental Science, 2022. IOP Publishing, 012170.

Liu, Z. & Karimi, I. a. J. E. 2020. Gas turbine performance prediction via machine learning. 192, 116627.

Lobo, J. L., Ballesteros, I., Oregi, I. & Del Ser, J. J. a. P. A. 2019. Real-time Electrical Power Prediction in a Combined Cycle Power Plant.

Mardin Abdullah, A., Ghassan Akram, Q. & Abbas Mohamad, A. 2024. Ocular Disease Classification Using Different Kinds of Machine Learning Algorithms. Zanco Journal of Pure and Applied Sciences, 36, 25-34.

Mohapatra, A. & Sanjay, P. 2014. Comparative Analysis of Inlet Air Cooling Techniques Integrated To Cooled Gas Turbine Plant. Journal of the Energy Institute, 88.

Pachauri, N. 2024. An emission predictive system for CO and NOx from gas turbine based on ensemble machine learning approach. Fuel, 366, 131421.

Qi, Z., Feng, Y., Wang, S. & Li, C. 2025. Enhancing hydropower generation Predictions: A comprehensive study of XGBoost and Support Vector Regression models with advanced optimization techniques. Ain Shams Engineering Journal, 16, 103206.

Rawal, K. & Ahmad, A. A Comparative Analysis of Supervised Machine Learning Algorithms for Electricity Demand Forecasting. 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T), 1-3 March 2022 2022. 1-6.

Santarisi, N. S. & Faouri, S. S. J. E.-E. J. O. E. T. 2021. Prediction of combined cycle power plant electrical output power using machine learning regression algorithms. 6, 114.

Shaik, N. B., Jongkittinarukorn, K. & Bingi, K. 2024. XGBoost based enhanced predictive model for handling missing input parameters: A case study on gas turbine. Case Studies in Chemical and Environmental Engineering, 10, 100775.

Sharifzadeh, M., Sikinioti-Lock, A., Shah, N. J. R. & Reviews, S. E. 2019. Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression. 108, 513-538.

Solyali, D. 2020. A Comparative Analysis of Machine Learning Approaches for Short-/Long-Term Electricity Load Forecasting in Cyprus. Sustainability [Online], 12.

Wankhede, A. & Ghate, V. Prediction of power of a combine gas and steam turbine using artificial neural network. 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2018. IEEE, 1103-1108.

Yi, Q., Xiong, H. & Wang, D. J. E. 2023. Predicting Power Generation from a Combined Cycle Power Plant Using Transformer Encoders with DNN. 12, 2431.