Pertanika J. Sci. & Technol. 33 (S4): 81 - 97 (2025)

PERTANIKA
JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Development of Digital Twin Data-driven Modelling for Gas Turbine Operation Behaviour

Balbir Shah Mohd Irwan Shah, Asnor Juraiza Ishak*, Mohd Khair Hassan and Nor Mohd Haziq Norsahperi

Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ABSTRACT

Digital twins have recently gained attention as digital solutions in "Energy 4.0" that will reshape the future of the power generation industry toward the digital era. It is supported by the rapid advancement of data connectivity and computational power to intensify the potential of digital twin technology in addressing the energy trilemma. The energy trilemma has been identified as a global challenge to transform the power generation industry landscape to be more efficient and competitive. Digital twins have been identified as a key enabler to address the impacts of this global challenge on power plants due to several factors such as ageing, performance degradation, and high operating costs. This study will evaluate the concept of the digital twin approach by developing the gas turbine digital twin to provide future insights into operational performance and optimisation. The gas turbine digital twin model is developed through a cutting-edge data-driven approach, utilising an artificial neural network (ANN) to deliver superior performance in advanced monitoring applications. The digital twin model is constructed structurally in four steps: process identification, data collection, pre-processing, and developing the digital twin plant model. The gas turbine operating parameters are analysed for critical parameter verification to emulate the gas turbine operation behaviour environment. The best deep learning structure for data-driven methods is identified based on a lower Mean Squared Error (MSE) and an average error of less than 0.5% of the predicted value. The findings indicate that the digital twin data-driven modelling can be applied to future advanced monitoring of gas turbines in the power generation industry.

Keywords: Artificial Neural Network (ANN), data-driven, digital twin, gas turbine, power plant, predictive modelling

ARTICLE INFO

Article history:
Received: 01 April 2024
Accepted: 10 April 2025
Published: 10 June 2025

DOI: https://doi.org/10.47836/pjst.33.S4.05

E-mail address:
balbirshahphd@gmail.com (Balbir Shah Mohd Irwan Shah)
asnorji@upm.edu.my (Asnor Juraiza Ishak)
khair@upm.edu.my (Mohd Khair Hassan)
nmhaziq@upm.edu.my (Nor Mohd Haziq Norsahperi)
*Corresponding author

e-ISSN: 2231-8526

INTRODUCTION

Digital twins play a significant role in addressing the Energy Trilemma, which has been identified as the global challenge of balancing energy security, environmental sustainability, and affordability. In the context of Industry 4.0, digital twins bring benefits to future power plantsby optimizing operational and maintenance costs. A digital twin represents a real thing, system, process, environment, or entity that exists digitally. Physical data from various sources, including sensors, IoT devices, and other data streams, is gathered and integrated to produce a digital twin model. This replica can be utilized for analysis, modeling, monitoring, and control purposes to better understand and manage the physical counterpart. In addition, the digital twin can be developed to perform advanced applications such as operating behaviour emulation, anomaly detection (Wu et al., 2023), and performance optimization (Rahman et al., 2011).

In conjunction with Malaysia's electricity supply industry (MESI 2.0) initiatives, the government is committed to addressing the energy trilemma by introducing new enhanced dispatch arrangement (NEDA) rules to transform the industry into a more efficient and productive one. Under NEDA, generators are allowed to compete based on lower variable operating rates (VOR) and heat rates rather than a fixed rate. In the worst-case scenario, a power plant cannot dispatch due to process disruptions and equipment failures. This will likely result in a prolonged shutdown to investigate the root cause and perform corrective action. This condition will have a significant impact on national grid security and sustainability due to several factors, such as aging, performance degradation (Meher-Homji et al., 2001), restricted loading operations (Castillo et al., 2021), and high operating costs (Zhong et al., 2023). Therefore, the digital twin has been identified as a key enabler in the future to address the impacts of electricity liberalization on thermal power plant operation and performance. This will enhance competition in generation dispatch and result in more competitive energy prices.

The digital twin model is developed through four steps: process identification, data collection, data pre-processing, and digital twin plant modeling. Initially, the study examines critical gas turbine parameters for verification to develop a digital twin data-driven model for advanced monitoring. The digital twin is then modeled using a deep learning approach to predict gas turbine operating behaviour for anomaly detection and performance analysis applications. Finally, the best deep learning structure for data-driven methods is identified. The outcome will revolutionize the power plant industry for various Industry 4.0 applications.

Literature Review

This study found that adopting digital twins has led to the discovery of five categories of power generation in the distribution of digital twin applications. The publication trend indicates that the adoption of the digital twin is dominated by coal-fired and gas turbine plant studies at 22.2%, followed by nuclear power and renewable energy at 18.5%, respectively. Meanwhile, the least studied renewable energy and cogeneration plants are 14.8% and 3.7%, respectively. Besides that, another finding indicates that a coal-fired power plant publication is focused on the study of components and processes at the same 50% rate. However, current studies on gas turbine plants predominantly focus on individual components rather than overall processes (Shah et al., 2024). As a result, this study will focus on developing a gas turbine digital twin, with a particular emphasis on process areas. The details of the digital twin study under the gas turbine application are shown in Table 1. The combined cycle power plant is a viable future alternative due to its high efficiency and capacity to address the energy trilemma while also realising the COP26 initiatives by 2030. The coexistence of this plant with renewable energy in the smart grid will help to rebalance the energy system between security and sustainability at an optimum cost. Hence, the gas turbine prospect can be explored to merge with renewable energy plants for smart grid implementation in the future. In addition, the emissions can be reduced further by introducing the new combustion technology that can generate energy from natural gas co-fired with clean fuel sources such as hydrogen, ammonia, and biomass.

Table1
The summary of the digital twin study under the gas turbine application

Authors	**Objective**	**Focus**	**Method**	**Category**
Tsoutsanis et al., 2020	Health monitoring for transient operation	Parts	Machine Learning	Component
Marwaha & Kohn, 2019	Predictive maintenance	Compressor	Physic Model	Component
Ren et al., 2017	Life prediction	Rolling bearing	Deep Learning	Component
Polyakov et al., 2020	Failure identification and prediction	Component	ANN	Component
Nikolaev et al.,2019	Condition monitoring for maintenance	Flame tubes	ANN, Physics Model	Component
Malozemov et al., 2019	Performance monitoring and improvement	Diesel engine	Physic Model	Component
Dawes et al., 2019	Life Cycle Modelling for MRO	Turbine blade	Geometry Model	Component

METHODOLOGY

This study aims to develop a digital twin model for gas turbine power plant applications to simulate the dynamic operational behavior of the system. The scope will encompass critical aspects of gas turbine components and overall processes. Figure 1 provides an overview of the methodology applied in this study. An artificial neural network (ANN) has been selected as the data-driven approach for developing the gas turbine digital twin. Moreover, this study will incorporate an optimizer algorithm to enhance prediction accuracy and improve the model's performance. This approach is expected to outperform previous ANN models, which often required extended iteration times due to the complexity of hidden layers and neurons.

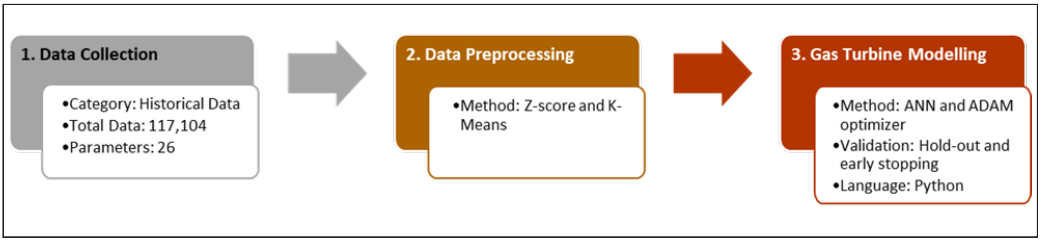


Figure 1. The summary of the methodology

Data Collection

The gas turbine's historical operating data is archived in a distributed control system (DCS) used to operate and control the power plant. The information is accumulated from a real power plant and recorded in CSV file format. The information was collected and categorised into four groups of periods that have been logged every minute for historical data collection. The Jupyter Notebook performs data analytics to provide an overview of data conditions and gas turbine operating profiles. The dataset contains a total of 117,104 data points and 26 parameters, comprising 24 inputs and 2 outputs.

The dataset is used to analyze variations, such as mean, median, mode, standard deviation, and maximum and minimum values of each parameter. Statistical analysis is essential for classifying values into different ranges and identifying the aberrant parameters and factors that impact performance and optimisation. Histogram plots are used to reveal outlier values during the operation of the gas turbine. In addition, the relationship between the parameters is presented using a correlation matrix. This matrix provides a quantitative indicator of the linear relationships between variables, helping to spot patterns and links within the data (Wagavkar, 2023).

Data Pre-processing

The dataset must be cleaned during the data pre-processing phase. Cleaning can standardise and structure data to maintain uniformity throughout a dataset and adding in fixing errors and inconsistencies to improve the accuracy of the data. This is crucial for data integrity and avoiding misleading and wrong analysis results. The dataset collected from the physical model is limited by the fact that anomalies and outliers can exist in the raw dataset for a variety of reasons, including data input mistakes, sensor failures, and unusual conditions. The actual and predicted input and output values may differ because of the presence of abnormal data points. Therefore, the dataset for the load in the generated output will be filtered as values higher than or equal to 125 MW will be used in this study.

The outlier removal methods, Z-score and K-means, will be applied in this study. In terms of standard deviations, the Z-score measures how much a data point deviates from the mean of a distribution. Outliers are often removed using Z-scores by defining a threshold over which data values are deemed outliers. Conversely, K-means is an iterative technique for grouping data that divides a dataset into a preset number of groups. To remove outliers, it seeks to minimise the within-cluster sum of squares, which indicates the compactness and similarity of data points within each cluster. The Z-scores and K-means techniques are utilised to eliminate the outliers from the dataset. Furthermore, a boxplot is constructed to visualise the data distribution before and after the data-cleaning process. Any value outside of the acceptable range has the probability of being an anomalous data point. Data points not within the acceptable range are considered an outliers. However, other statistical analyses and domain knowledge are still required to make informed decisions about the presence and handling of outliers.

Gas Turbine Modelling

The selected approach for a data-driven task is an artificial neural network (ANN) that utilizes backpropagation, a robust algorithm for supervised learning. Backpropagation is noteworthy for its ability to determine discrepancies between the network's predictions and actual outcomes. It works by switching the direction errors move through the network's layers. This lets neuron-associated weights be precisely adjusted to improve predictions. For the practical implementation of this ANN model, Jupyter Notebook has been chosen due to its compatibility with Python, which is renowned for rapid execution. Additionally, Jupyter Notebook supports a vast array of machine learning and deep learning libraries, making it an excellent tool for this study. Artificial neural networks are powerful tool within the domains of artificial intelligence and machine learning, adept at handling tasks like pattern recognition, classification, regression, and decision-making. The versatility of ANNs allows them to classify and train on data, enabling the generalisation of strategies for data interpretation during evaluations.

In this study, the ANN model is developed using the Keras library and structured with input, hidden, and output layers suitable for gas turbine applications. The model ensures data consistency by normalising input data to have a zero mean and unit variance. The composition of ANNs includes neurons, input layers, hidden layers, and output layers, arranged in a layered architecture. The hidden layer settings consists of three because a large amount of data requires attentive analysis to reduce errors and losses, which enhances the accuracy of the prediction results and achieves the desired outcomes. The connections between neurons are associated with weights. These weights determine the strength of the connection and are adjusted during training. The initial weights of connections are frequently chosen at random using initialisation methods. Each neuron in a layer calculates the weighted sum of its inputs, adds a bias term, and then uses an activation function. The bias factor increases the weighted sum before the activation function is used. The network output is compared to the desired target following a forward pass. The gradients of this comparison for the network weights are computed via backpropagation. The learning rate determines the step size of these weight updates. This procedure continues until the network converges or achieves a suitable performance level. The ANN will generate predictions on the original data after training.

Load and Pre-process Data

The StandardScaler from the Python scikit-learn Library is applied to load and pre-process datasets. The StandardScaler standardizes data on the output and input features by eliminating the mean and scaling to unit variance. Each input and output column is scaled using this approach by subtracting the mean and dividing by the standard deviation. The scaled features are then returned and provided with input and output values. The data is converted to a similar scale by standardizing the input attributes and output target. In addition, MinMaxScaler transform the data to a fixed range, usually between 0 and 1. Normalizing the variables can help alleviate these difficulties and boost the optimisation algorithm's convergence.

Model Training and Evaluation

The pre-processed data is split into training and testing sets. The model trains over multiple epochs, updating its parameters after processing batches of 32 samples. With a 70% training and 30% validation split, performance is monitored using holdout validation and early stopping. This approach is particularly effective, given the dataset of 117,104 data points and the hardware limitations that restrict extended training times for complex models. Early stopping enhances the holdout method by stopping training once the model reaches its best performance on the validation set, preventing overfitting. It halts training once optimal performance is achieved on the validation set, reducing overfitting. Meanwhile, the scheduler technique is used to automatically adjust the learning rate based on validation loss, reducing it when improvement stagnates to help the model overcome local minima. This improves convergence without manual intervention, thereby optimizing the training process. After compiling the model with a loss function, optimizer, and metrics, a sequential model with three dense layers is created. The model is tested on a new dataset to minimize bias, and performance metrics are used to fine-tune hyperparameters.

Structure of the ANN Model

Figure 2 illustrates the structure of the Artificial Neural Network (ANN), consisting of three hidden layers. The first hidden layer utilizes the rectified linear unit (ReLU) activation function and contains 256 neurons. This layer's input data is characterized by 24 features. The second hidden layer also employs the ReLU activation function and consists of 128 neurons. The third and final layer uses a linear activation function with two output units. The activation functions are essential for introducing non-linearity to the model, allowing the network to capture complex patterns. The output layer uses a linear activation function because the problem is a regression task, where the model predicts continuous output values. The neuron counts in the hidden layers are powers of two, which optimize memory allocation and enhance performance.

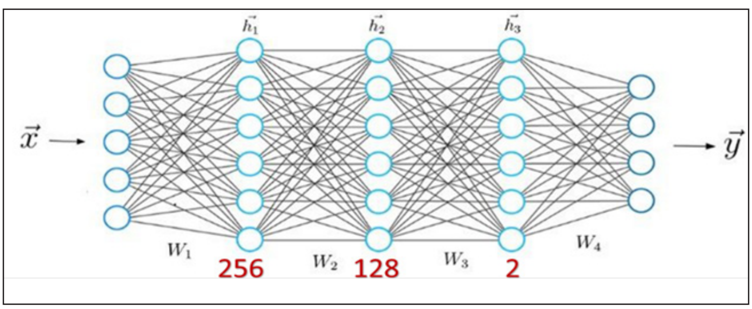


Figure 2. ANN development structure

RESULT AND DISCUSSION

The gas turbine digital twin was modeled using artificial neural networks (ANN) and achieved optimal performance by integrating the ADAM (adaptive moment estimation) optimizer. A thorough performance evaluation focused on critical metrics such as mean absolute error (MAE), root mean square error (RMSE), learning curves, gas turbine loss, test loss, regression values, prediction accuracy, and overall digital twin performance. This analysis clearly highlighted the transformative impact of data cleaning by comparing model performance before and after the process. The digital twin model's performance was rigorously tested, starting with a modest architecture of 256 neurons in the first layer and 128 in the second. The model was progressively scaled to larger configurations, with the most effective setup found to be 2048 neurons in the first layer and 1024 in the second, optimized by the ADAM algorithm. This architecture was identified as the ideal structure for enhancing gas turbine performance. After completing the training and validation phases, the model's performance on the test dataset was carefully evaluated. As outlined in Table 2, both MAE and RMSE show significant reductions following data cleaning, indicating a substantial improvement in the model's predictive accuracy. The decrease in these metrics confirms that removing outliers effectively minimizes discrepancies between predicted and actual values. This demonstrates the model's robustness and reliability, setting a new benchmark for data-driven gas turbine performance analysis modelling.

Table 2
Overall data-driven model performance results

Model performance		
No.	**Performance Indicator**	**Results**
1	MAE	4.3758e-04
2	MSE	2.3419e-04
3	RMSE	0.1004
4	Test Accuracy	0.0101
5	R^2	1.00

Furthermore, the test accuracy improves after data cleaning compared to before. The increase in the number of neurons and the addition of hidden layer, as well as the use of the ADAM optimizer for prediction, result in lower test loss and higher test accuracy. The regression value (R^2) indicates the relationship between the output and input variables. The negative coefficient indicates that the input and output are negatively correlated, meaning they tend to move in opposite directions; however, the regression value after data cleaning is positively correlated, implying that it tends to move in the same direction. An R^2 score of 1.0 indicates a perfect fit.

The loss function curves are plotted to display the training and validation accuracy and loss over time. Figure 3 shows the MSE loss function curve, illustrating the neural network model's training loss across epochs. Lower loss values during both training and validation indicate improved model performance. After data pre-processing, the validation loss is less than the training loss, signaling that overfitting has not occurred. Data pre-processing led to better overall performance, as it removed disturbances present in the previously trained and tested datasets. Before data cleaning, the validation line indicated overfitting, where the model performed well on training data but poorly on test data. This overfitting would have negatively impacted the gas turbine's output performance. Addressing this issue improved the model's robustness and generalizability.

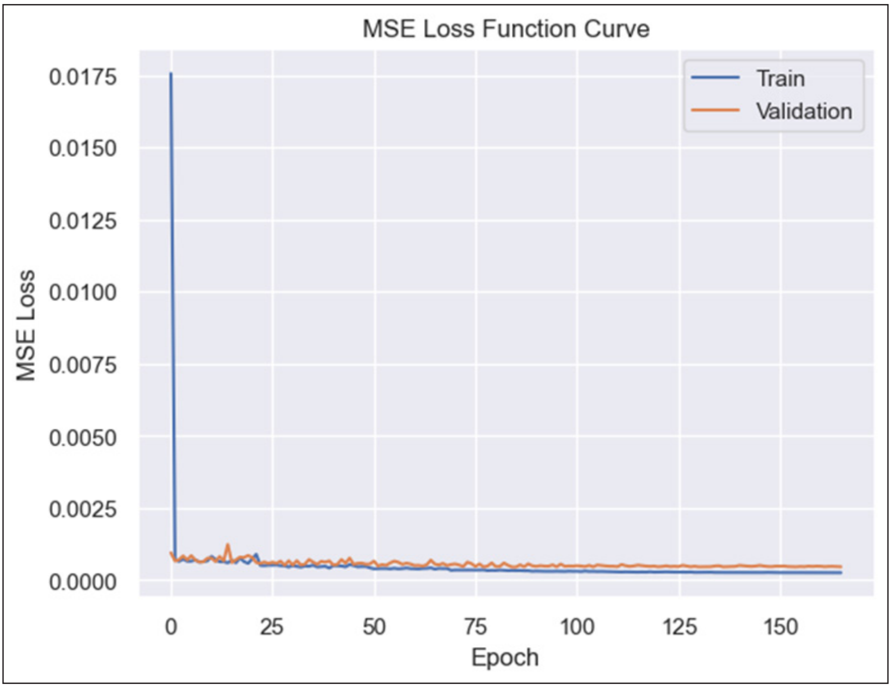


Figure 3. MSE loss function curve

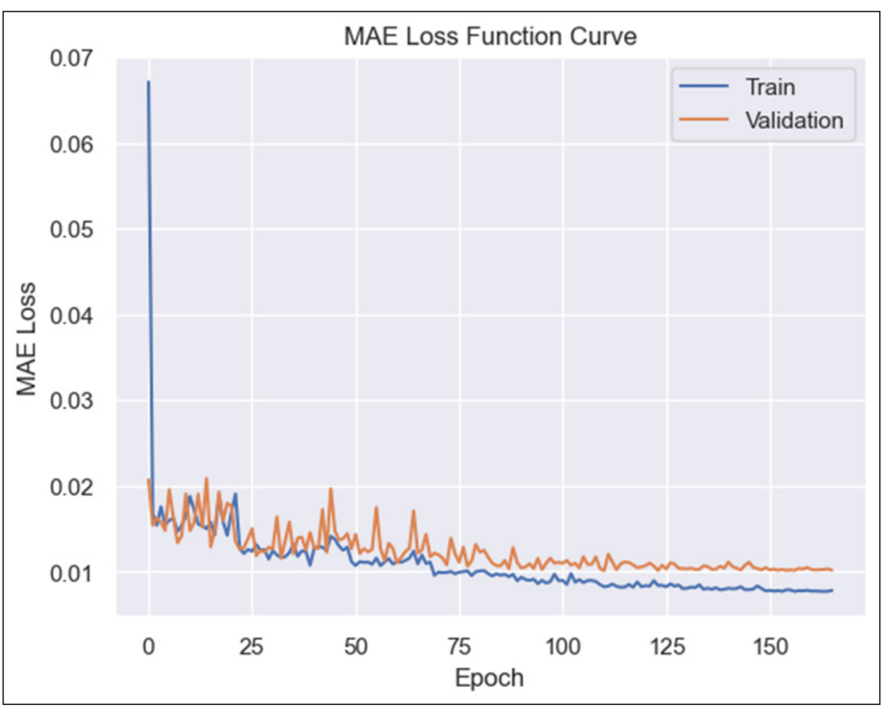


Figure 4. MAE loss function curve

The MAE is a key metric to evaluate the model's ability to generalize and predict new datasets. As shown in Figure 4, the MAE decreases with increasing epoch values, signifying that the model is progressively learning and improving. MAE validation is continuously monitored throughout the training process to ensure strong performance on unseen data, helping to avoid problems like overfitting or underfitting. An early stopping mechanism is applied during training, halting the process when validation MAE ceases to improve, thus preventing overfitting. The steady decline in errors and the minimal gap between training and validation MAE reflect stable validation, indicating that the model generalizes effectively without significant increases in validation error over time.

Separate graphs are created for each input and output variable to compare predicted values to actual values before data cleaning. According to Figure 5, the input predicted value overlaps with the actual value, implying that the predicted value is identical to the actual value. If the predicted value and actual value do not overlap and are far apart, it indicates that the actual value is not at the predicted value and is influenced by external disturbances in the gas turbine. After data cleaning, the predicted and actual values for all parameters are nearly identical. In contrast, the dataset before data cleaning contains a few data points that are not identical. This is because outliers and disturbances affect dataset learning and prediction accuracy. After data cleaning, the location of the data points changes because the outliers that affected the gas turbine's performance were eliminated.

Figure 6 illustrates the comparison between predicted and actual output values, showing a strong correlation as represented by a diagonal line. After data cleaning, the predicted values for gas turbine load and efficiency align closely with the actual values, highlighting the improved performance of the model. Prior to data cleaning, some predicted values exhibited deviations from the actual outputs, though the spread remained relatively narrow. A narrower spread signifies higher prediction accuracy, while a wider spread indicates greater variability in model performance. The dense clustering of predicted values after cleaning demonstrates high forecasting accuracy, whereas more dispersed plots before cleaning suggest increased variability. Overall, the improved alignment in the cleaned data underscores the enhanced accuracy and effectiveness of the data-driven model in predicting gas turbine load and efficiency.

The normal operational behaviour is illustrated by comparing real and predicted values of input and output parameters at specific loads, including 125 MW, 150 MW, 180 MW, 200 MW, and 220 MW. Table 3 presents samples of these values, along with the error percentages and cumulative error percentages. The results show that predictions after data cleaning are more accurate, with lower error percentages compared to those before data cleaning. The accuracy of predictions before data cleaning is compromised by outliers, which affect the model's ability to predict desired values accurately.

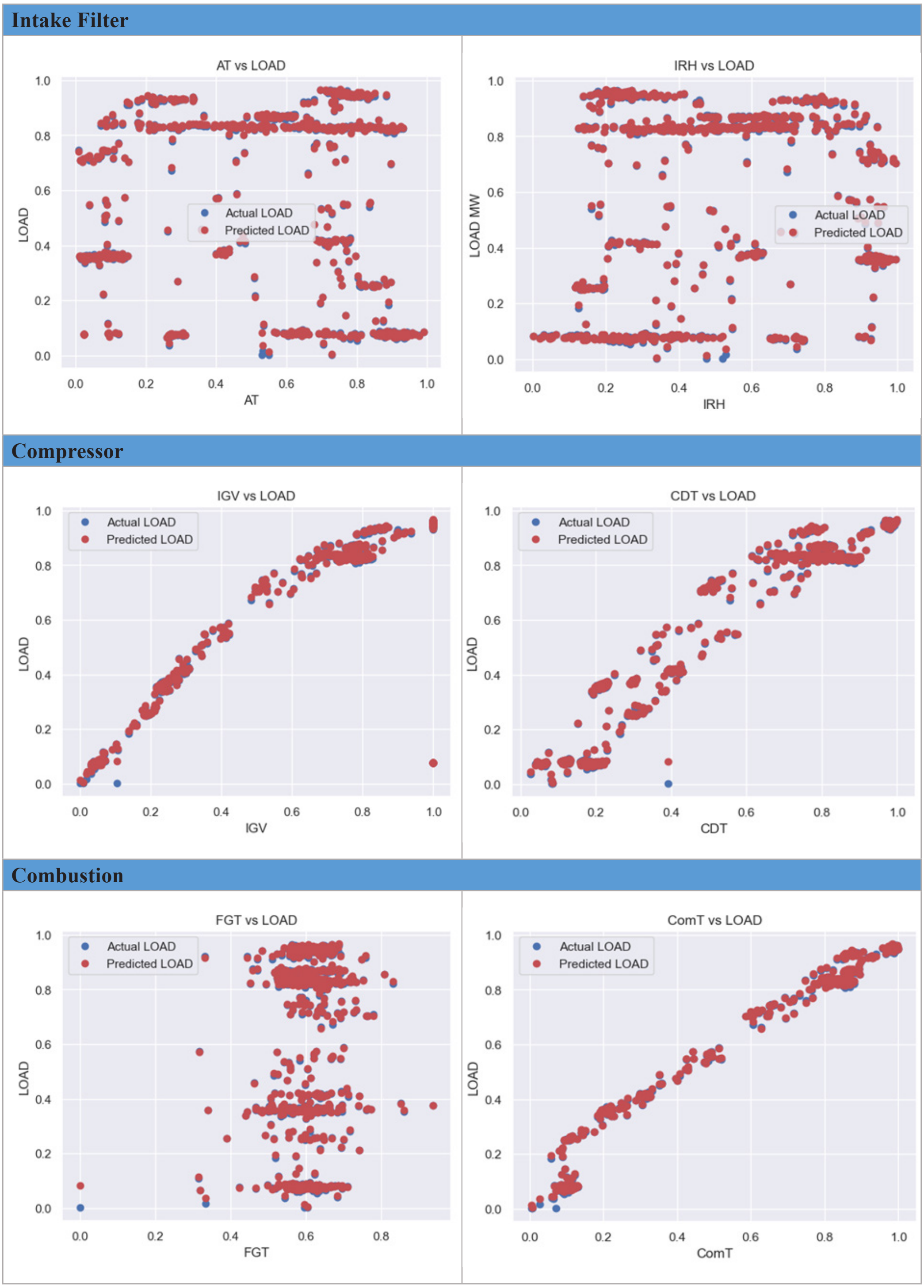


Figure 5. The comparison between the actual and predicted values of the input parameters

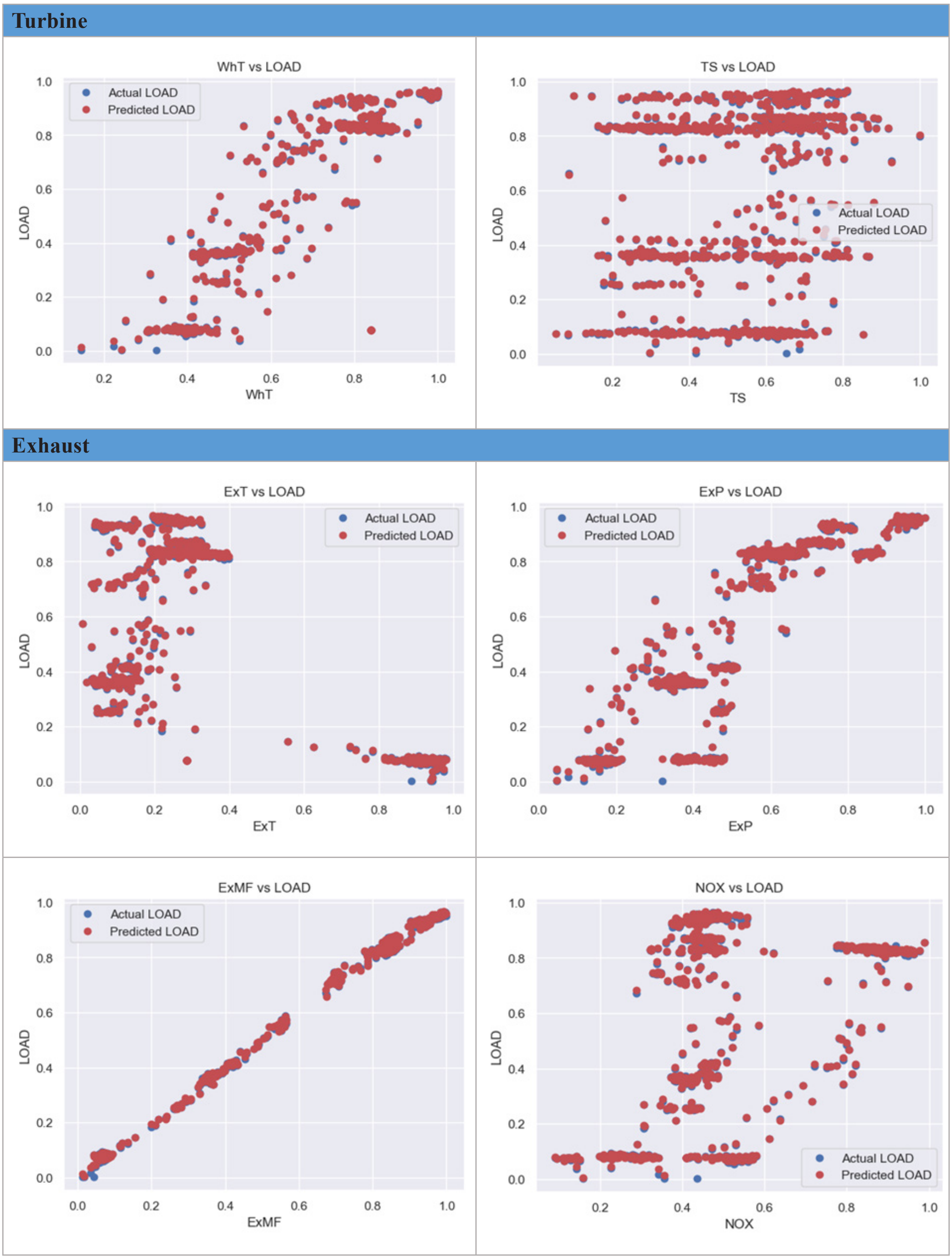
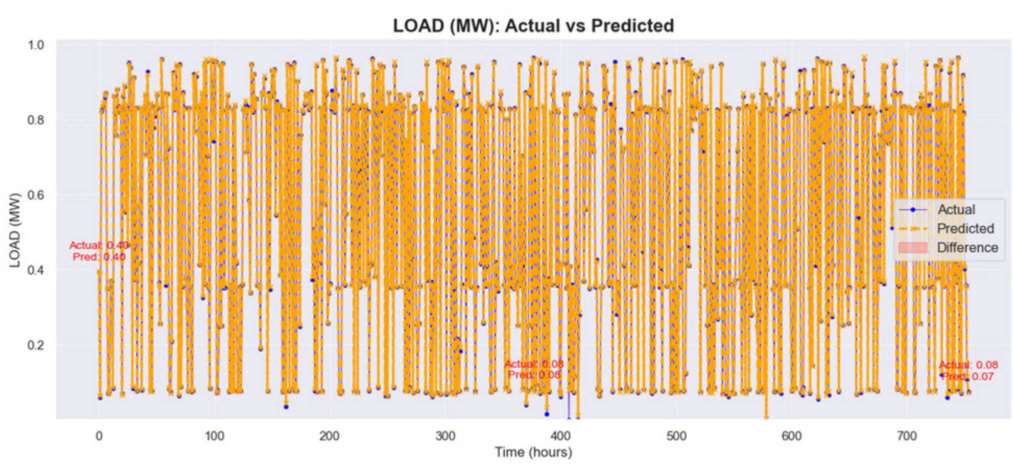


Figure 5 (continue)

Load



Efficiency

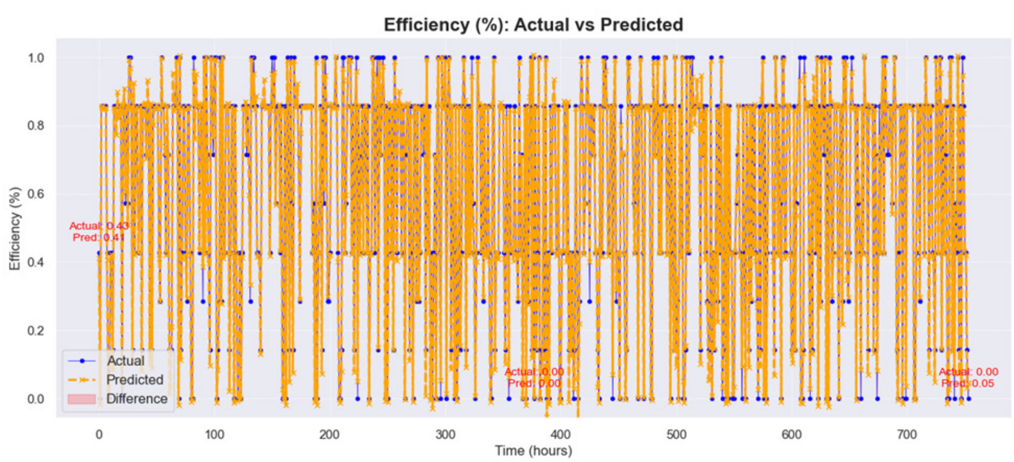


Figure 6. The comparison between the actual and predicted values of the output parameters

Table 3
Predicted gas turbine operating parameters at 150 MW

Actual vs Predicted						
No	**Operating Parameter**		**Real Value**	**Prediction Value**	**Error**	**% Error**
1	Intake Filter	Ambient Temperature (degC)	25.84	25.9260	-0.0860	0.3328
		Pressure (Bar)	1.01	1.0055	0.0045	0.4455
		Relative Humidity	79.4	79.0366	0.3634	0.4577
		Mass Flow (m3/hr)	1615.92	1612.5670	3.3530	0.2075

Table 3 *(continue)*

Actual vs Predicted						
No	**Operating Parameter**		**Real Value**	**Prediction Value**	**Error**	**% Error**
2	Compressor	Inlet Guide Vanes (DEG)	55.18	54.6566	0.5234	0.9485
		Discharge Temperature degC)	360.55	360.9410	-0.3910	0.1084
		Discharge Pressure (Bar)	10.31	10.3290	-0.0190	0.1843
3	Combustion	Fuel Gas Flow (t/hr)	35.71	35.7924	-0.0824	0.2307
		Fuel Gas Interstage Pressure (Bar)	26.84	26.8396	0.0004	0.0015
		Fuel Gas Temperature (degC)	185.24	185.3047	-0.0647	0.0349
		Speed Ratio Valve (%)	48.97	48.9502	0.0198	0.0404
		Fuel Stroke Reference (%)	50.77	50.7708	-0.0008	0.0016
		Gas Fuel LHV (Btu/scf)	935.93	934.3998	1.5302	0.1635
		Firing Temperature (degC)	1234.39	1235.3215	-0.9315	0.0755
4	Turbine	Wheelspace Temperature (degC)	442.88	442.4450	0.4350	0.0982
		Turbine Speed (RPM)	2999.37	2999.5833	-0.2133	0.0071
5	Exhaust	Temperature (degC)	633.73	633.6862	0.0438	0.0069
		Pressure (bar)	27.8	27.9665	-0.1665	0.5989
		Exhaust Mass Flow (t/hr)	1655.1	1657.1015	-2.0015	0.1209
		Nitrogen Oxides (Nox) (ppm)	20.04	20.0313	0.0087	0.0434
		Sulfur Dioxide (SO2) (ppm)	0.11	0.1117	-0.0017	1.5455
		Carbon Monoxide (CO) ppm)	1.97	1.9772	-0.0072	0.3655
		Carbon Dioxide (CO2) vol%)	4.08	4.0968	-0.0168	0.4118
		Oxygen (O2) (vol%)	13.62	13.6015	0.0185	0.1358
6	Generated Output	Load (MW)	150.16	149.8041	0.3559	0.2370
		Efficiency (%)	30	30.0840	-0.0840	0.2800
Total percentage of error (%)						7.0840

Table 4 comprehensively analyses predictive error percentages for gas turbine loadings ranging from 125 MW to 220 MW in a cumulative measure. This measure is derived from the sum of the error percentages for 26 different parameters. Notably, the determined average for these predictive errors is 0.23%. This result demonstrates the effectiveness of the predictive models in action.

Table 4
Overall performance at various loads

Load	Error %	Ave Error %
125MW	7.8198	0.3008
150MW	7.0840	0.2725
180MW	5.1220	0.1970
200MW	6.4324	0.2474
220MW	4.3914	0.1689
Average error %		0.2373

CONCLUSION

The findings of this study mark a significant advancement in the realm of gas turbine performance analytics. This study rigorously built a digital twin model of a gas turbine using artificial neural networks (ANN) and achieved a commendable average error percentage of approximately 0.23% across various operating conditions. This exceptional precision indicates a highly successful model development process, demonstrating the effectiveness of the ANN methodology in capturing the complex performance dynamics of gas turbines. The ANN hyperparameters were optimized throughout the study, resulting in significant improvements in the model's predictive accuracy, as evidenced by substantial decreases in key performance metrics such as mean absolute error (MAE) and root mean squared error (RMSE). Such enhancements reinforce the model's robustness and highlight the diligent efforts to improve the analytical instrument. However, the study's insights are particularly valuable for gas turbine operations at load levels exceeding 50% capacity. The complexity of turbine behavior at lower loads presents additional challenges, making the current model less effective in these volatile conditions. Future research will address these challenges by enhancing the digital twin model's ability to perform under low-load conditions and incorporating predictions from a thermodynamic cycle perspective.

ACKNOWLEDGEMENT

The authors express their heartfelt gratitude to the management team of Tuanku Ja'afar Power Station and TNB Power Generation Sdn. Bhd. (Genco) for their invaluable support in this study. Their generosity in granting permission and assistance in gathering critical real-time data unequivocally contributed to the scholarly success of this research work. This collaboration shows the positive outcomes that can be achieved through industry-academic collaboration.

REFERENCES

Castillo, I. G., Loboda, I., & Pérez Ruiz, J. L. (2021). Data-driven models for gas turbine online diagnosis. *Machines*, *9*(12), 372. https://doi.org/10.3390/machines9120372

Dawes, W. N., Meah, N., Kudryavtsev, A., Evans, R., Hunt, M., & Tiller, P. (2019, January 7-11). Digital geometry to support a gas turbine digital twin. [Paper presentation] *AIAA Scitech 2019 Forum*, San Diego, California. https://doi.org/10.2514/6.2019-1715

Malozemov, A. A., Solomonenko, M. V., & Malozemov, G. A. (2019). Numerical simulation of power plants with reciprocating engines using Modelica language. In *2019 International Russian Automation Conference (RusAutoCon)* (pp. 1-5). IEEE. https://doi.org/10.1109/RUSAUTOCON.2019.8867801

Marwaha, G., & Kohn, J. (2019). Predictive maintenance of gas turbine air inlet systems for enhanced profitability as a function of environmental conditions. In *Abu Dhabi International Petroleum Exhibition and Conference* (pp. SPE-197814). SPE. https://doi.org/10.2118/197814-ms

Meher-Homji, C. B., Chaker, M. A., & Motiwala, H. M. (2001). Gas turbine performance deterioration. In *Proceedings of the 30th turbomachinery symposium*. Texas A&M University. Turbomachinery Laboratories.

Nikolaev, S., Belov, S., Gusev, M., & Uzhinsky, I. (2019). Hybrid data-driven and physics-based modelling for prescriptive maintenance of gas-turbine power plant. In C. Fortin, L. Rivest, A. Bernard & A. Bouras (Eds.) *Product Lifecycle Management in the Digital Twin Era* (pp. 379–388). Springer. https://doi.org/10.1007/978-3-030-42250-9_36

Polyakov, R., Paholkin, E., Kudryavcev, I., & Krupenin, N. (2020). Improving the safety of power plants by developing a digital twin and an expert system for adaptive-predictive analysis of the operability of gas turbine units. *Turbo Expo: Power for Land, Sea, and Air, 84157*, Article V006T09A001. https://doi.org/10.1115/GT2020-14217

Rahman, M. M., Ibrahim, T. K., & Abdalla, A. N. (2011). Thermodynamic performance analysis of gas-turbine power-plant. *International Journal of the Physical Sciences*, *6*(14), 3539-3550. https://doi.org/10.5897/IJPS11.272

Ren, L., Cui, J., Sun, Y., & Cheng, X. (2017). Multi-bearing remaining useful life collaborative prediction: A deep learning approach. *Journal of Manufacturing Systems*, *43*, 248–256. https://doi.org/10.1016/j.jmsy.2017.02.013

Shah, B. S. M. I., Ishak, A. J., Hassan, M. K., & Norsahperi, N. M. H. (2024). Energy 4.0: Challenges and enablers of digital twin application in power plant. *ELEKTRIKA-Journal of Electrical Engineering*, *23*(1), 103-124. https://doi.org/http://dx.doi.org/10.11113/elektrika.v23n1.487

Tsoutsanis, E., Hamadache, M., & Dixon, R. (2020). Real-time diagnostic method of gas turbines operating under transient conditions in hybrid power plants. *Journal of Engineering for Gas Turbines and Power*, *142*(10), Article 101002. https://doi.org/10.1115/1.4048340

Wagavkar, S. (2023). I*ntroduction to the correlation matrix.* Builtin. https://builtin.com/data-science/correlation-matrix

Wu, Z. Y., Chew, A., Meng, X., Cai, J., Pok, J., Kalfarisi, R., Lai, K. C., Hew, S. F., & Wong, J. J. (2023). High fidelity digital twin-based anomaly detection and localization for smart water grid operation management. *Sustainable Cities and Society*, *91*, Article 104446. https://doi.org/10.1016/j.scs.2023.104446

Zhong, D., Xia, Z., Zhu, Y., & Duan, J. (2023). Overview of predictive maintenance based on digital twin technology. *Heliyon, 9*(4), Article e14534. https://doi.org/10.1016/j.heliyon.2023.e14534