Dynamic Operations Conditions of Industrial Gas Turbine Emissions Monitoring: XGBoost-DT-SVM Hybrid Stacked Ensemble Model

Rudy Winarto
School of Interdisciplinary Management and Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
7032241011@student.its.ac.id

Wiwik Anggraeni
Department of Information Systems
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
wiwik@its.ac.id

Mauridhi Hery Purnomo
Department of Electrical Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
hery@ee.its.ac.id

***Abstract*—Accurate emission prediction is essential for maintaining regulatory compliance and optimizing industrial gas turbine performance. Conventional Predictive Emissions Monitoring System (PEMS) often exhibit reduced accuracy under dynamic and nonlinear operating conditions. This study introduces a hybrid stacked ensemble framework that integrates XGBoost, Decision Tree, and Support Vector Machine base learners with an ElasticNet meta-learner to enhance robustness, accuracy, and computational efficiency. Using cleaned operational datasets, the model predicts four key pollutants: CO, CO_2, SO_2, and NO. Experimental results show that the stacked ensemble consistently outperforms individual learners, achieving an average coefficient of determination (R^2) of approximately 0.95 and MAPE values below 5%, meeting industrial acceptance criteria. The ElasticNet meta-learner further improves generalization and reduces error variability, particularly under transient and fluctuating load conditions. The findings highlight the XGBoost-DT-SVM ensemble as a cost-effective and regulation-aligned alternative to Continuous Emissions Monitoring System (CEMS), advancing intelligent and adaptive data-driven PEMS architectures.**

***Index Terms*—CEMS, DT, ensemble learning, gas turbine, hybrid machine learning, PEMS, SVM, XGBoost**

I. Introduction

A. Background

Global efforts to address climate change have intensified the need to reduce emissions from industrial sectors, including oil and gas, where gas turbines remain major sources of CO_2, NO_x, and SO_2. While essential for power generation and processing, turbine combustion produces highly variable emissions across steady and transient operating regimes. Accurate monitoring is therefore critical for optimizing combustion efficiency and ensuring regulatory compliance.

Conventional CEMS provide direct measurements but are costly, inflexible, and maintenance-intensive, driving interest in data-driven Predictive Emission Monitoring Systems (PEMS) as a more economical alternative. Advances in machine learning have enabled models capable of capturing nonlinear relationships between operating parameters and emissions; however, balancing accuracy, interpretability, and computational efficiency remains challenging. This study develops a hybrid stacked ensemble designed to maintain robust performance under non-stationary industrial turbine conditions, supporting more intelligent and sustainable emission-prediction frameworks.

B. Problem Statement

Gas turbine emissions such as CO and NO_x require strict compliance with environmental standards [1]. Traditional CEMS, though accurate, are costly and lack flexibility for dynamic industrial environments [2]. Many machine learning-based PEMS models do not segment data by operational regime [3], reducing reliability during transient conditions and limiting interpretability of regime-specific performance [6]. Few studies incorporate adaptive or regime-aware modeling approaches [7]. Integrating operational state awareness, steady and transient has shown substantial benefits for capturing nonlinear, regime-dependent emission behavior [8], underscoring the need for more robust and practical PEMS solutions.

II. Literature Review

A. Related Works

Prior research has introduced hybrid machine learning models such as ANN-SVM, GRNN-GA [9], and LSTM-HGB for emissions prediction, but few have considered operational regime segmentation. Studies [10] [11] highlight that model accuracy changed significantly during load changes or startup transitions. Recent ensemble methods (XGBoost, LightGBM) improve prediction accuracy but rarely address computational cost or adaptability. The proposed work bridges this gap through performance-balanced, regime-aware modeling.

B. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting, also known as XGBoost is a popular machine learning algorithm due to its efficiency, accuracy, and scalability in predictive modeling [12]. The term "gradient" refers to the process of sequentially decreasing error from the previous model to the next. On the other hand, the term "boosting" refers to combining several weak learners, each improving on the errors of the previous one [13].

C. Decision Tree (DT)

Decision Tree machine learning [14], is a popular method used in both classification and regression tasks. The primary principle of a DT is to divide the dataset into subgroups and use the most significant features to produce the greatest difference in the tree. This method has been widely adopted due to its interpretability, robustness, and ability to handle complex datasets. Each node in the tree represents a feature, and each branch represents a decision rule based on that feature.

D. Support Vector Machine (SVM)

The Support Vector Machine is a powerful supervised machine learning technique that is usually used for classification and regression by leveraging its strengths in handling non-linear data [15]. Support Vector Machine (SVM) works with the five key steps e.g. selection of features and target variables, plotting the data (2D) to visualize the data points and separating the classes, linear or non-linear kernel selection based on data complexity, training the model to optimize the hyperplane by maximizing the margin between classes, and making a prediction to classify new data based on the learned hyperplane.

E. Ensemble Machine Learning

Ensemble learning refers to algorithms that combine the predictions from two or more base models to improve overall performance and generalization. It is a widely adopted machine learning approach that enhances predictive accuracy by aggregating diverse model outputs. According to [16], ensemble methods are typically categorized into three main types: bagging, stacking, and boosting. Each method has distinct mechanisms and use cases, and understanding their principles is essential for developing robust predictive models.

Stacking Ensemble or stacked generalization, combines multiple heterogeneous base learners and uses a meta-learner to synthesize their outputs, unlike bagging or boosting, which often rely on homogeneous models. Stacking leverages the strengths of diverse algorithms, for instance, combining XGBoost, DT, and SVM. The meta-learner learns to weight or integrate predictions from base models to achieve superior performance. Stacking is particularly effective when base learners capture complementary data patterns, thereby improving predictive robustness and generalization.

F. Gas Turbine Operational Regimes

Gas turbine operational data are clustered into three primary regimes, startup, steady-state, and transient using load percentage or turbine speed thresholds. During this phase, the system operates under unstable thermodynamic conditions, characterized by frequent fluctuations in flow rate, temperature, and pressure. The turbine typically operates at low load or near-idle (approximately 20-30% of its rated output), resulting in incomplete combustion and sub-optimal air-fuel mixing. Consequently, this phase is associated with elevated levels of carbon monoxide emissions due to insufficient oxidation. Although the startup phase is relatively short, it produces significant emission spikes that critically influence overall environmental performance and model calibration accuracy.

The steady-state phase corresponds to the turbine's optimal operational regime, where process parameters remain stable and efficiency reaches its peak. In this phase, the turbine typically operates at 70-100% of its rated load, achieving thermodynamic equilibrium and consistent combustion performance. Emissions exhibit predictable behavior, with relatively stable concentrations of carbon monoxide, nitrogen oxides, and sulfur dioxide. These steady conditions enable near-maximum combustion efficiency and minimal fluctuation in exhaust parameters. Because of its operational stability and data consistency, the steady-state phase provides an ideal basis for training and validating baseline PEMS models, ensuring accurate and reproducible performance across industrial applications.

The transient phase is characterized by rapid load ramp-ups, shutdown sequences, or quick cycling between operating states. This regime represents the most complex and challenging condition to model due to its highly dynamic and non-linear behavior. Sudden variations in temperature and pressure often lead to spikes in nitrogen oxide emissions, reflecting delayed combustion responses and transient mixing inefficiencies. Emission patterns during this phase often exhibit time lags relative to operational adjustments, leading to predictive instability and increased modeling uncertainty. Although the transient data are typically noisier than steady-state operation, they are critical for capturing the actual dynamics of industrial gas turbine behavior and ensuring that predictive models remain robust under realistic operating conditions.

G. Research Originality and Contribution

It addresses unexamined gaps in emissions monitoring, such as the need to incorporate environmental factors into gas turbine emissions models and to classify data into startup, steady-state, and transient phases. The originality is further supported by the focus on creating a balanced model that combines multiple performance metrics, a feature uncommon in traditional emissions monitoring studies.

For academics, it advances the field of emissions prediction by proposing a new framework that balances multiple factors (accuracy and computational efficiency) and leverages the strengths of different machine learning models in an ensemble. For the industry, the outcomes of this research could provide more reliable, cost-effective emissions-monitoring solutions better suited for real-time applications.

III. Methodology

A. Data Acquisition and Preprocessing

For the study, a compressor unit powered by a gas turbine and operating at a low pressure is chosen. The suction pressure is 30 psig, the discharge pressure is 210 psig, and the gas flow rate is designed at a maximum of 100 MMscfd. A gas turbine drives the low-pressure compression system with a capacity of 39,520 horsepower (*hp*). Historical and real-time data were collected from multi-year gas turbine operations. Preprocessing steps include outlier removal, normalization, and feature selection based on correlation and mutual information scores [17].

Parameters such as load expressed in head of energy or as a percentage of capacity, fuel flowrate, turbine inlet temperature, compressor discharge pressure, exhaust gas temperature, air-fuel ratio, and ambient conditions provide a comprehensive view of the turbine's performance under various scenarios. By analyzing these variables collectively, clustering techniques can effectively differentiate between regimes or conditions such as startup, steady-state, and transient operations, each of which visualizes unique emission characteristics and system dynamics.

B. Collection of the Data

The used dataset is Process Historian Data with the duration of January 1, 2021 to December 31, 2024 with 5 minutes sampling rate. A total dataset of 101,120 for each 15 input and 4 output variables. In this research the threshold used for startup if LOAD $=<$30%, steady state if LOAD $>$30%, and transient of changes of the LOAD $>$10% in 5 minutes.

C. Optimum Data Split

The ratio of training to test data should be determined by the model's complexity and the number of parameters in the dataset. In case there are three sets, the optimal dataset split sizes should be as per [18].

$$\text{Split ratio} = p : \sqrt{p} : (\sqrt{p} + 1) \tag{1}$$

where p is the number of parameters. In this research, the parameters are 19, and the split ratio is 68.3%:15.7%:16.1%.

D. Hybrid Stacked Ensemble

The proposed model integrates XGBoost, DT, and SVM as heterogeneous base learners, each contributing unique algorithmic strengths. The DT offers high interpretability and captures rule-based hierarchical relationships between process parameters and emission outputs. The SVM introduces strong generalization capability through non-linear kernel mapping, effectively handling complex boundaries and multi-dimensional feature interactions. Meanwhile, XGBoost, a gradient-boosting ensemble algorithm, provides computational efficiency and robustness by combining multiple weak learners and applying regularization to prevent overfitting.

This combination is referred to as a hybrid ensemble because it integrates base models derived from different learning paradigms, tree-based (DT, XGBoost) and support vector-based (SVM), to leverage their complementary characteristics. Tree-based models excel at handling feature interactions, categorical variables, and interpretability, while SVMs offer greater precision in high-dimensional, non-linear regions of the feature space. By merging these algorithmic approaches, the hybrid structure achieves a balance between accuracy, generalization, and explainability.

The outputs of the three base learners are combined by a meta-learner using a weighted averaging strategy. This stacking mechanism ensures that the meta-learner learns the optimal weights for each base learner's predictions, thereby improving overall performance and stability across different operational regimes. As a result, the hybrid ensemble not only enhances predictive accuracy but also maintains interpretability and computational efficiency suitable for industrial deployment.

E. Selected Model Algorithms Performance Assurance

To ensure the long-term effectiveness of the selected machine learning algorithm under continuously changing operational conditions, it is essential to develop a resilient, adaptive model that can respond to variations in process behavior or data distribution. One approach to achieving this is through mechanisms for continuously validating model performance, enabling the system to detect changes in input characteristics or predictive accuracy.

In deployment, this requires real-time error monitoring using metrics such as MAPE and RMSE, which are compared against actual measured outcomes. Additionally, it is critical to evaluate model performance across varying operational regimes, such as transient, steady-state, and startup conditions, and, where applicable, to incorporate measures of predictive confidence (e.g., prediction intervals).

When error metrics exceed predefined thresholds, the system should automatically initiate model retraining to maintain predictive reliability and support robust performance in dynamic industrial environments. In accordance with industrial application, the MAPE of all predictive models was constrained to a maximum threshold of 5%, ensuring compliance and operational reliability.

IV. Result

A. Segmentation of Gas Turbine Dynamic Operations Phase

Startup phase: this phase involves ignition and early ramp-up. The turbine is unstable, with frequent changes in flow rate, temperature, and pressure. Low load or near-idle conditions are typical (e.g., $<$ 20-30% capacity).

Full-load or steady-state phase: the turbine runs at optimal efficiency with stable operating parameters. Emissions are more predictable, and the system reaches thermal equilibrium. Operation occurs at stable load and emission levels.

Transient phase: characterized by rapid load ramps or drops. This regime is the most complex and typically the hardest to model. It includes load ramp-ups, shutdown sequences, or quick cycling between states.

B. SEM Model Performance

The results in Table I summarize the performance of the hybrid SEM model compared with its base learners. Among the models, the Meta Learner achieves the lowest RMSE (1.576) and MSE (2.484), indicating superior predictive precision and generalization capability. The DT and XGBoost models also

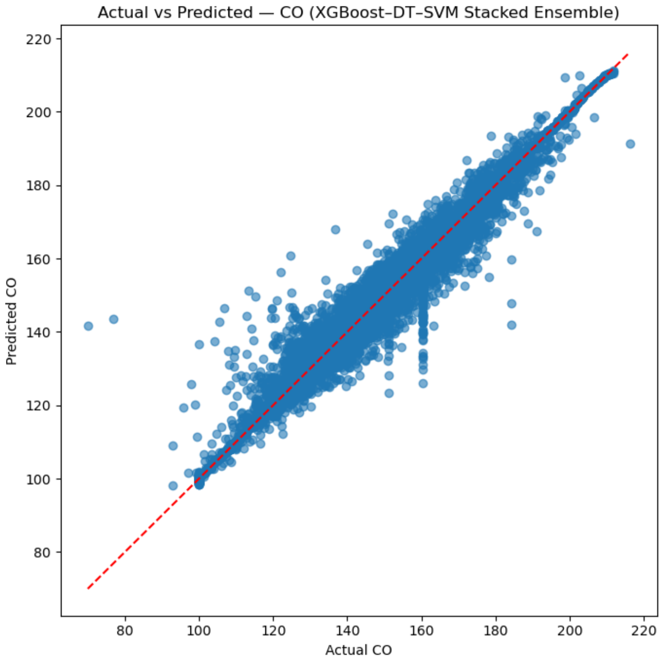

Fig. 1. SEM model CO actual-predicted plot

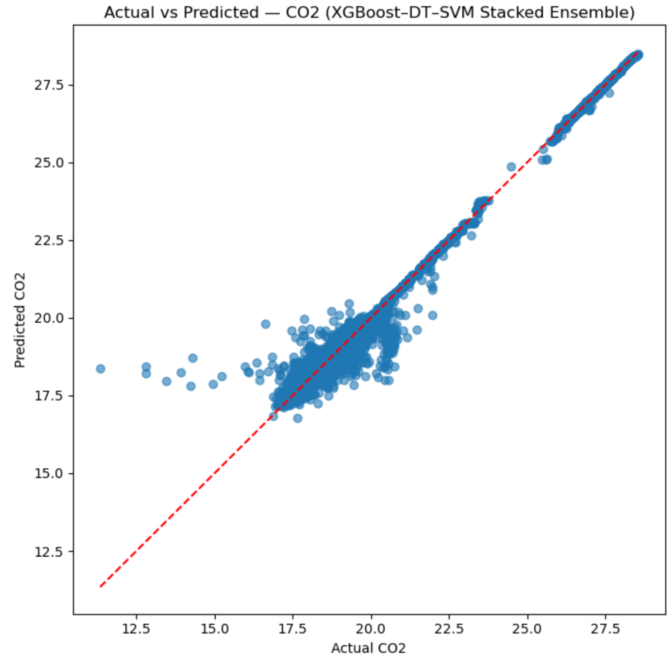

Fig. 2. SEM model CO_2 actual-predicted plot

perform competitively with similar error levels, whereas the SVM shows the highest errors and lowest R^2 value (0.813), reflecting weaker predictive reliability. Overall, the stacked ensemble effectively integrates the strengths of individual learners, providing the best balance between accuracy and robustness for NO emission prediction.

TABLE I
SEM MODEL PERFORMANCE METRICS

Target	Model	RMSE	MAPE	R^2
CO	XGBoost	1.623771	1.418681	0.973689
CO	DT	3.706039	1.139995	0.975010
CO	SVM	11.061794	5.392608	0.777366
CO	Meta-learner	3.607409	1.307361	0.976323
CO_2	XGBoost	0.250904	0.606337	0.993696
CO_2	DT	0.241270	0.468250	0.994171
CO_2	SVM	0.849401	3.001372	0.927756
CO_2	Meta-learner	0.235239	0.546052	0.994459
NO	XGBoost	1,623715	3.070087	0.945388
NO	DT	1.628333	2.591217	0.945076
NO	SVM	3.003000	8.655562	0.813198
NO	Meta-learner	1.576150	2.860436	0.948540
SO_2	XGBoost	1.602449	9.127876	0.964480
SO_2	DT	1.259190	3.802677	0.978069
SO_2	SVM	4.367073	125.314233	0.736211
SO_2	Meta-learner	1.320990	7.077160	0.975863

The R^2 comparison plots collectively demonstrate that the proposed XGBoost-DT-SVM model achieves strong predictive performance across all emission parameters (CO, CO_2, SO_2, and NO). For CO and CO_2, both the DT and meta-learner attain near-perfect coefficients of determination ($R^2 > 0.97$), indicating that the ensemble effectively captures both linear and nonlinear relationships with minimal residual error. The meta-learner slightly outperforms the base models, confirming the benefit of combining diverse learners to improve generalization. For SO_2 and NO, the DT and meta-learner again deliver superior accuracy ($R^2 > 0.94$), while SVM performs comparatively weaker due to its sensitivity to high variance and complex operational dynamics. Overall, the ensemble consistently enhances robustness and stability, minimizing overfitting while maintaining high interpretability and computational efficiency, making it well-suited for real-time industrial PEMS deployment.

The Figure 1 present scatter plot of actual versus predicted CO values, demonstrating a strong positive linear correlation, with data points closely aligned along the 45° reference line. This alignment indicates that the stacked ensemble effectively captures the underlying emission behavior of CO with minimal systematic bias. The dense clustering of points near the diagonal suggests high predictive accuracy and stable generalization across different operating conditions. Only a small dispersion is observed at higher CO values, likely reflecting transient load variations or sensor noise. Overall, the model demonstrates reliable estimation and confirms the robustness of the hybrid ensemble framework for accurately predicting CO emissions.

The Figure 2 exhibits a strong linear correlation, with most data points tightly clustered along the 45° reference line. This close alignment demonstrates that the stacked ensemble accurately captures the underlying CO_2 emission behavior across varying operational conditions. The minimal deviation at lower concentrations indicates a slight underestimation during transient states, likely due to short-term fluctuations in combustion dynamics. Overall, the model delivers excellent

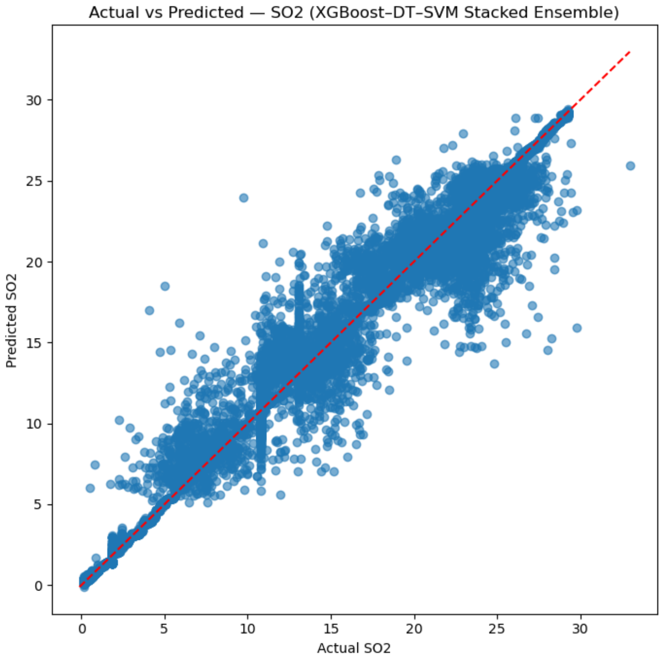

Fig. 3. SEM model SO_2 actual-predicted plot

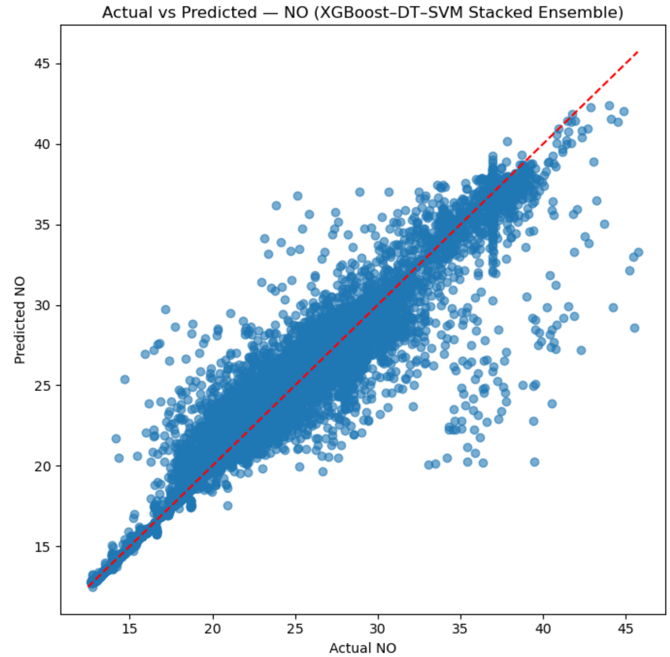

Fig. 4. SEM model NO actual-predicted plot

predictive accuracy, confirming the ensemble's robustness and reliability in estimating CO_2 emissions.

The Figure 3 reveals a strong positive correlation, with most data points distributed closely along the 45° reference line. This indicates that the stacked ensemble effectively captures the nonlinear relationships governing SO_2 emission behavior. While a few isolated points appear at higher concentrations, likely due to transient operating conditions or measurement uncertainty, the overall trend remains consistent and accurate. The close alignment between predicted and actual values confirms the ensemble model's ability to generalize well across diverse turbine conditions, resulting in reliable SO_2 emission estimates.

The Figure 4 illustrates the correlation between the actual and predicted NO_x emission values produced by the SEM model. A strong linear alignment of data points along the 45° reference line indicates that the ensemble model effectively learns the nonlinear relationships between input parameters and emissions across varying turbine operating conditions. The tight clustering of points around the diagonal demonstrates excellent agreement between observed and predicted values, confirming the model's high accuracy and minimal systematic bias. Only a limited number of data points deviate significantly from the reference line, mainly in higher regions, suggesting occasional underestimation or overestimation under transient or high-load combustion scenarios, conditions that typically exhibit rapid variations in temperature and fuel–air mixing. The relatively uniform spread along the regression line also suggests that the model maintains stable predictive performance across the entire emission range, without significant skewness.

The use of XGBoost, Decision Tree, and SVM as base learners enables the stacked ensemble to capture both localized nonlinear behavior and global emission trends efficiently. This synergistic combination improves generalization and reduces variance compared to individual models. Overall, the figure demonstrates that the hybrid approach achieves robust predictive accuracy and strong generalization for NO_x emission monitoring, making it suitable for real-time PEMS deployment in industrial gas turbine environments where accuracy, computational efficiency, and regulatory compliance are critical operational requirements.

V. Discussion

A. Model Performance and Comparative Insights

Tree-based learners, especially XGBoost and Decision Tree, delivered the highest accuracy and stability, consistent with prior findings on their strength in handling structured industrial data. Incorporating SVM in the ensemble improved the capture of fine nonlinear patterns, particularly during transient load shifts, while the ElasticNet meta-learner effectively balanced these complementary behaviors to reduce variance. Overall, the heterogeneous stacking strategy provided stronger generalization than any single or homogeneous model.

B. Regime-Dependent Emissions Characteristics

Segmenting the dataset into startup, steady-state, and transient regimes enabled clearer interpretation of emission behavior. Startup conditions showed elevated CO and SO_2 due to incomplete combustion, while transient phases introduced noise and lag effects, challenging all base models. Despite this, the stacked ensemble remained accurate across regimes, outperforming traditional PEMS approaches that assume uniform operating conditions. Steady-state periods produced the lowest residuals, confirming their value for calibration.

C. Comparison with Previous Studies

Earlier hybrid models (e.g., ANN-SVM, GRNN-GA) offer high accuracy but suffer from limited interpretability and heavy computational demand. The proposed ensemble preserves transparency while achieving comparable or better accuracy at lower runtime cost. Compared with deep architectures such as LSTM, the stacked model achieves similar R^2 scores with significantly reduced training requirements. These results extend findings by [19] and [20], showing that optimized stacking can surpass both standalone and sequential hybrid learners for multi-emission prediction.

VI. Conclusions and Recommendations

A. Conclusions

The XGBoost-DT-SVM stacked ensemble, combined with an ElasticNet meta-learner, offers a robust and efficient PEMS solution for industrial gas turbines. The model achieved MAPE $< 5\%$ and $R^2 \approx 0.95$, outperforming all base learners in accuracy and stability. Its regime-aware design effectively captures nonlinear and transient emission dynamics, positioning it as a viable, cost-effective alternative to conventional CEMS for real-time monitoring and compliance assurance.

B. Recommendations and Future Works

Industrial deployment should integrate the stacked ensemble into a full PEMS workflow with sensor validation, real-time preprocessing, and periodic recalibration. Coordination between data scientists and field engineers is essential to align retraining schedules with maintenance cycles. Regulatory bodies should consider certifying validated data-driven PEMS frameworks when they consistently meet compliance thresholds.

Future research should incorporate incremental or online learning to address operational drift, sensor degradation, and fuel variability. Integrating explainability together with uncertainty quantification would also improve traceability and auditability of model outputs, strengthen operator confidence, and support regulatory acceptance by providing defensible evidence of key emission drivers and quantified prediction reliability across operating regimes.

References

[1] Pachauri, N. (2024). An Emission Predictive System for CO and NOx from Gas Turbine based on Ensemble Machine Learning Approach. Fuel, 366, 131421. https://doi.org/10.1016/j.fuel.2024.131421.

[2] Chien, T. W., Chu, H., Hsu, W. C., Tseng, T. K., Hsu, C. H., and Chen, K. Y. (2003). A Feasibility Study on the Predictive Emission Monitoring System Applied to the Hsinta Power Plant of Taiwan Power Company. Journal of the Air and Waste Management Association, 53(8), 1022–1028. https://doi.org/10.1080/10473289.2003.10466241.

[3] Dalal, A. S., Sultanova, N., Jayabalan, M., and Mustafina, J. (2023). Gas Turbine – CO and NOx Emission Data Analysis With Predictive Modeling Using ML/AI Approaches. 2023 16th International Conference on Developments in ESystems Engineering (DeSE), 100–104. https://doi.org/10.1109/DeSE60595.2023.10469322.

[4] dos Santos Coelho, L., Hultmann Ayala, H. V., and Cocco Mariani, V. (2024). CO and NOx Emissions Prediction in Gas Turbine using a Novel Modeling Pipeline based on the Combination of Deep Forest Regressor and Feature Engineering. Fuel, 355, 129366. https://doi.org/10.1016/j.fuel.2023.129366.

[5] Tuttle, J. F., Blackburn, L. D., Andersson, K., and Powell, K. M. (2021). A Systematic Comparison of Machine Learning Methods for Modeling of Dynamic Processes Applied to Combustion Emission Rate Modeling. Applied Energy, 292. https://doi.org/10.1016/j.apenergy.2021.116886.

[6] Wood, D. A. (2023). Long-term Atmospheric Pollutant Emissions from a Combined Cycle Gas Turbine: Trend Monitoring and Prediction Applying Machine Learning. Fuel, 343, 127722. https://doi.org/10.1016/j.fuel.2023.127722.

[7] Vanderhaegen Laborelec Linkebeek, E., Michaël Deneve Laborelec Linkebeek, B., Hannes Laget Laborelec Linkebeek, B., Nathalie Faniel Laborelec Linkebeek, B., and Jan Mertens Laborelec Linkebeek, B. (2010). GT2010-22899 Predictive Emissions Monitoring using a Continouosly Updating Neural Network AMS Automated Measuring System CEMS Continuous Emission Measurement System ELV Emission Limit Value LCPD Large Combustion Plants Directive PEMS Predictive Emission Monitoring System QAL Quality Assurance Level 2. Power for Land.

[8] Morshed-Bozorgdel, A., Kadkhodazadeh, M., Valikhan Anaraki, M., and Farzin, S. (2022). A Novel Framework Based on the Stacking Ensemble Machine Learning (SEML) Method: Application in Wind Speed Modeling. Atmosphere, 13(5), 758. https://doi.org/10.3390/atmos13050758.

[9] Winarto, R., Purnomo, M. H., and Anggraeni, W. (2024). Predicting CO_2 Emission of Gas Turbine Using Hybrid GRNN-GA Machine Learning Approach. 2024 4th International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS), 264–269. https://doi.org/10.1109/ICE3IS62977.2024.10775309.

[10] Bartolini, C. M., Caresana, F., Comodi, G., Pelagalli, L., Renzi, M., and Vagni, S. (2011). Application of artificial neural networks to micro gas turbines. Energy Conversion and Management, 52(1), 781–788. https://doi.org/10.1016/j.enconman.2010.08.003.

[11] Hashmi, M. B., Lemma, T. A., Ahsan, S., and Rahman, S. (2021). Transient Behavior in Variable Geometry Industrial Gas Turbines: A Comprehensive Overview of Pertinent Modeling Techniques. Entropy, 23(2), 250. https://doi.org/10.3390/e23020250.

[12] Feng, M., Wang, X., Zhao, Z., Jiang, C., Xiong, J., and Zhang, N. (2024). Enhanced Heart Attack Prediction Using eXtreme Gradient Boosting. Journal of Theory and Practice of Engineering Science, 4(04), 9–16. https://doi.org/10.53469/jtpes.2024.04(04).02.

[13] Chen, T., and Guestrin, C. (2016). XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785.

[14] Quinlan. (1986). Induction of Decision Trees. In Machine Learning (Vol. 1).

[15] Frenica, A., Lindawati, L., Lindawati, L., Soim, S., and Soim, S. (2023). Implementasi Algoritma Support Vector Machine (SVM) untuk Deteksi Banjir. INOVTEK Polbeng - Seri Informatika, 8(2), 291. https://doi.org/10.35314/isi.v8i2.3443.

[16] Yang, Y., Li, Y., Cheng, L., and Yang, S. (2024). Short-Term Wind Power Prediction Based on a Modified Stacking Ensemble Learning Algorithm. Sustainability (Switzerland), 16(14). https://doi.org/10.3390/su16145960.

[17] Winarto, R., Purnomo, M. H., and Anggraeni, W. (2025). Enhancing Predictive Emissions Monitoring Performance: Data Preprocessing for XGBoost-Based Model Algorithm. 2025 17th International Conference on Knowledge and Smart Technology (KST), 278–283. https://doi.org/10.1109/KST65016.2025.11003353.

[18] Joseph, V. R. (2022). Optimal ratio for data splitting. Statistical Analysis and Data Mining: The ASA Data Science Journal, 15(4), 531–538. https://doi.org/10.1002/sam.11583.

[19] Tuttle, J. F., Blackburn, L. D., Andersson, K., and Powell, K. M. (2021). A Systematic Comparison of Machine Learning Methods for Modeling of Dynamic Processes Applied to Combustion Emission Rate Modeling. Applied Energy, 292. https://doi.org/10.1016/j.apenergy.2021.116886.

[20] dos Santos Coelho, L., Hultmann Ayala, H. V., and Cocco Mariani, V. (2024). CO and NOx Emissions Prediction in Gas Turbine using a Novel Modeling Pipeline based on the Combination of Deep Forest Regressor and Feature Engineering. Fuel, 355, 129366. https://doi.org/10.1016/j.fuel.2023.129366.