Research Paper

Data-driven Analysis and Optimization of Combined Cycle Power Plants using Machine Learning Models

Seksaf Hiba Dikra[1], Amjad Gawanmeh[2,*], & Khelif Rabia[1]

[1]Department of Mechanical Engineering, Badji Mokhtar-Annaba University, 17 hassen chaouche, Annaba 23000, Algeria
[2]College of Engineering and IT, University of Dubai, Academic City, Emirates Road, 14143, Dubai, UAE

*Corresponding author: agawanmeh@ud.ac.ae

Abstract

The current global energy demand relies more on Combined Cycle Power Plants (CCPPs) for their high efficiency and reduced environmental footprint. However, the performance of these plants is very sensitive to several environment parameters including temperature, pressure, humidity, and exhaust vacuum. This paper is intended to use machine learning (ML) approach to model and optimize CCPP energy production based on these factors. The proposed method uses a dataset with hourly environmental measurements, to provide detailed analysis using ML techniques including Random Forests and Neural Networks to identify any potential nonlinear relationships and predict energy output. The results showed that ambient temperature has the most significant influence on energy production, followed by vacuum, pressure, and humidity. In addition, this paper also highlighted optimal environmental conditions that maximize energy output, which can help and support power plant operators in optimizing their operation factors. In summary, the recommendations and outcomes of this paper provide necessary steps for integrating advanced ML techniques into CCPP operations, enhancing both efficiency and sustainability.

Keywords: *combined cycle power plants (ccpps); energy production; machine learning (ml); optimization; predictive modelling.*

Introduction

The global energy sector has witnessed a significant shift towards combined cycle power plants (CCPPs) due to their high efficiency and reduced environmental impact (Arrieta & Lora, 2005). These plants integrate gas and steam turbines in a symbiotic manner, where the exhaust heat from the gas turbine, typically wasted in simple cycle operations, is harnessed to power a steam turbine, thereby substantially improving overall efficiency (Bagnasco et al., 1998). This dual-cycle approach has demonstrated efficiency rates of up to 60%, marking a considerable improvement over conventional single-cycle plants that typically achieve 35-40% efficiency (Fantozzi & Desideri, 1998). However, the performance of CCPPs is highly influenced by several ambient conditions, including temperature, atmospheric pressure, relative humidity, and exhaust vacuum (Wang et al., 2019). These environmental parameters affect the thermodynamic cycles of both gas and steam turbines, creating complex dependencies that impact net energy output, which cannot be directly measured or identified with traditional observation or methods. For instance, elevated ambient temperatures reduce air density, compromising gas turbine performance, while variations in atmospheric pressure affect combustion efficiency (Kotowicz & Brzęczek, 2018). Similarly, humidity levels influence the thermodynamic properties of air and combustion dynamics, while exhaust vacuum conditions directly impact steam cycle efficiency (Zwebek, 2002). Traditional approaches to analyzing CCPP performance have relied on empirical studies combined with thermodynamic calculations (Danish et al., 2023). However, these methods often struggle to capture the non-linear relationships and complex interactions between environmental variables and plant performance. The limitations of observation and conventional analytical methods results in gaps in the ability to accurately measure, predict and optimize CCPP operations under varying environmental conditions (Asghar et al., 2023). Hence, there is a need for advanced methods to capture and analyze such condition.

Recent advances in Artificial Intelligence (AI) and machine learning present promising solutions to these challenges. These technologies offer computational and statistical methods that can identify complex and hidden patterns within operational data, and can also model the relationship between environmental parameters and conditions and plant

operation performance (Xezonakis et al., 2024). In addition, machine learning algorithms can be useful in creating predictive models for power plant optimization based on the available operational data. This paper proposes a machine learning approach to model energy generation in CCPPs using the hourly environmental data to predict energy output under several operational conditions. The work is based on machine learning methods including linear regression, random forests and artificial neural networks, to identify the most effective modeling approach for accurate energy production predictions (Akdemir, 2016).

The economic implications of CCPP optimization are substantial. A 1% improvement in plant efficiency can translate to millions of dollars in annual fuel cost savings for a typical 400MW facility. With global CCPP capacity exceeding 1,000 GW and natural gas prices remaining volatile, even marginal performance improvements have significant economic impact. Additionally, increasingly stringent environmental regulations require plants to operate at peak efficiency to minimize emissions per MWh generated. This study addresses the critical need for data-driven optimization strategies that can help operators reduce fuel consumption by 2-5%, decrease maintenance costs through optimal operating conditions, and ensure compliance with emissions standards while maintaining grid reliability.

The remainder of this paper is organized as follows: Section 2 reviews related work in CCPP optimization using machine learning, identifying research gaps and motivating this research. Section 3 shows the objectives and research methodology. Section 4 describes the dataset and parameters used in the analysis. Section 5 presents the results and analysis of the machine learning models and their performance. Section 6 shows the optimization framework and practical implementation recommendations. Section 7 discusses the implications of the findings and addresses implementation challenges. Finally, Section 8 concludes the paper with a summary of key contributions and future research directions.

Related Work

Advances in Combined Cycle Power Plant Optimization and Performance Prediction Using Machine Learning

Research in combined cycle power plant (CCPP) optimization and performance prediction has evolved significantly over the past years, mainly with the integration of recent computational and automation methods such as artificial intelligence and machine learning approaches. This section examines the recent research in this domain.

The work in Arrieta and Lora (2005) addressed ambient temperature influence on CCPP performance, establishing correlations between environmental conditions and plant efficiency. Bagnasco et al. (1998) studied CCPP management during parallel and islanding operations, providing foundational understanding of plant dynamics. Authors in Fantozzi and Desideri (1998) applied artificial neural networks to power plant transient simulation, demonstrating the potential of AI in this domain. Recent research has increasingly focused on environmental parameters effects on CCPP performance. For instance, The study by Wang et al. (2019) analyzed CCPP performance with inlet air heating under partial load conditions, while Kotowicz and Brzęczek (2018) conducted a case study on improving modern CCPP efficiency. Authors in Zwebek (2002) provided analysis of combined cycle performance deterioration, identifying key factors affecting long-term plant efficiency.

Authors in Danish et al. (2023) developed an AI-coherent data-driven forecasting model, while the work in Asghar et al. (2023) focused on sustainable operations using AI-based power prediction. Xezonakis et al. (2024) demonstrated successful application of neural networks for output power estimation in both gas and combined cycle plants. Another contribution in Akdemir (2016) studied hourly power generation prediction using artificial neural networks. Saeed et al. (2023) implemented a recurrent neural network optimized by waterwheel plant algorithm for power output prediction. Authors in Yi et al. (2023) introduced transformer encoders with DNN for predicting power generation. On the other hand, authors in Andjelić et al. (2024) studied symbolic regression using genetic programming algorithms for electrical power output estimation.

Several researchers have conducted comparative analyses of different ML methods. For instance, the work in Bandić et al. (2020) compared random decision tree algorithms with ANFIS for power output prediction. In addition, authors in Kaewprapha et al. (2022) evaluated various machine learning approaches for estimating CCPP efficiency. Authors in Siddiqui et al. (2021) developed a comprehensive ML algorithm-based paradigm for power prediction. Recent approaches focused on practical implementation and real-time monitoring, where Khalid et al. (2021) developed a data-driven ML-based sensor selection approach for fault detection Sharma et al. (2023) created data-driven models for

power plants under cycling conditions, and finally Kabengele et al. (2022) demonstrated practical applications of ANN models in CCPP performance monitoring.

Researchers have also explored statistical and hybrid methodologies, such as the work in Dutta and Ghosh (2021) that presented a statistical approach to predicting electrical power output, Chary (2021) which focused on full load electrical power output prediction using various machine learning methods, and Manuel et al. (2024) which investigated the impact of AI on renewable energy utilization and visual power plant efficiency. In recent approaches, authors in Ntantis and Xezonakis (2024) optimized electric power prediction using innovative machine learning techniques, while authors in Ali (2021) explored prediction methods specifically for smart city applications. These studies suggest future directions in integrating advanced AI techniques with traditional power plant operations, and demonstrate the need for advanced methods to capture complex relationship between environmental parameters and plants performance. Table 1 below shows comparative analysis of recent significant works in the field.

Table 1   Summary of Recent Research in CCPP Optimization Using Machine Learning.

Study	Technique	Opt.	Vars	R^2	Factory Opt.	Key Features	Limitations
(Danish, Nazari, & Senjyu, 2023)	AI-coherent forecasting	✓	T,P,H,V	0.901	✓	Real-time prediction capability	High computational cost
(Asghar et al., 2023)	AI-based prediction	✓	T,P,H	0.892	✓	Sustainable operation focus	Limited environmental parameters
(Xezonakis, Samuel, & Enweremadu, 2024)	ANN	✗	T,P,H,V	0.887	✗	Combined gas and cycle prediction	No real-time adaptation
(Saeed et al., 2023)	RNN + Waterwheel	✓	T,P,H,V	0.923	✗	Advanced optimization algorithm	Complex implementation
(Yi, Xiong, & Wang, 2023)	Transformer + DNN	✓	T,P,H	0.912	✓	High accuracy prediction	Large dataset requirement
(Bandić, Hasičić, & Kevrić, 2020)	Random Decision Tree + ANFIS	✗	T,P	0.856	✗	Hybrid approach	Limited parameter scope
(Siddiqui et al., 2021)	ML-based paradigm	✓	T,P,H,V	0.898	✓	Comprehensive framework	Scalability issues
(Khalid, Hwang, & Kim, 2021)	ML-based sensor selection	✓	T,P,V	0.878	✓	Real-time fault detection	Limited to specific configurations
(Ntantis & Xezonakis, 2024)	Innovative ML	✓	T,P,H,V	0.915	✓	Power prediction optimization	High resource requirements

*T: Temperature, P: Pressure, H: Humidity, V: Vacuum, F: Fuel composition, L: Load factor. Opt. Optimization algorithm used.

Recent Trends and Gaps

The study of existing research work in the literature shows some critical gaps including:

1. Model Complexity vs. Performance: Some deep learning approaches like transformer encoders (Yi, Xiong, & Wang, 2023) and recurrent neural networks (Saeed et al., 2023) achieve higher accuracy, however, they demand high computational resources and very large datasets, which can be challenging to achieve and perform well in real time. In addition, traditional methods (Akdemir, 2016) remain simple and practical, but cannot capture hidden patterns and cannot provide predictive modes. This highlights the need for balanced approaches that optimize both accuracy and computational efficiency.
2. Environmental Parameter Integration: While recent studies (Wang et al., 2019; Danish et al., 2023) have improved parameter integration, most studies still focus on limited environmental variables. For instance, while (Arrieta & Lora, 2005) thoroughly examines temperature effects and (Asghar et al. 2023) considers multiple parameters, few studies address the complete set of variables including temperature, pressure, humidity, and exhaust vacuum simultaneously. This limits understanding of parameter interactions and their impact.
3. Real-time Optimization: While there is advancement in real-time monitoring systems (Khalid et al., 2021) and fault detection, there are many challenges in developing real time optimization solutions. Current implementations (Kabengele et al., 2022) cannot find the balance between real time response requirements and real time computational constraints. On the other hand, solutions (Ntantis & Xezonakis, 2024) that offers power prediction optimization, provide limited real-time adaptation capabilities.
4. Scalability and Adaptability: While some recent solutions (Bandić et al., 2020; Kaewprapha et al., 2022) perform well under specific conditions, they still have many practical challenges when adapted to different practical plant configurations. For instance, statistical approaches (Dutta & Ghosh, 2021) and hybrid methodologies lack flexibility and practical considerations in diverse practical operational scenarios, and lack the capability for accommodating varying plant configurations and operational conditions (Manuel et al., 2024).

These gaps highlight the need for improved methods, mainly in developing scalable solutions that can effectively balance computational efficiency with several parameter integrations as well as providing real-time optimization and operation capabilities. This study addresses this issue as described in the next section.

Objectives of the Study

While existing research has made progress in applying machine learning to CCPP optimization, several critical gaps remain unaddressed as described above. Many studies focus on single environmental variables or simple linear relationships, failing to capture the complex interactions between multiple environmental parameters. In addition, there is a lack of comparative analysis of plants performance in real operational settings. This study is intended to model the energy production of combined cycle power plants using an integrated AI approach. The first objective is developing a comprehensive environmental model that analyzes the simultaneous influence of multiple environmental variables (AT, AP, RH, V) on energy production, captures complex non-linear interactions between variables, particularly the critical AT-AP relationship, and quantifies the relative importance of each variable to inform operational priorities.

This study's unique contributions differ from prior work in several key aspects: (1) Simultaneous multi-parameter optimization: Unlike previous studies that examined parameters individually (Danish et al., 2023; Asghar et al., 2023) or in limited combinations (Bandić et al., 2020; Kaewprapha et al., 2022), this work develops a comprehensive model integrating all four critical environmental parameters with their interactions, (2) Practical implementation framework: While existing research focuses primarily on prediction accuracy, this study translates model outputs into specific operational recommendations with defined parameter ranges and real-time monitoring protocols, (3) Comparative model validation: The systematic evaluation of multiple ML approaches (Random Forests, Neural Networks, Linear Regression) with identical datasets provides practitioners with evidence-based algorithm selection criteria, and (4) Economic-operational integration: The optimization framework explicitly considers operational constraints alongside performance maximization, addressing the gap between theoretical optimal conditions and practical plant limitations.

The second objective is evaluating multiple models and conducting rigorous comparative analysis using both RMSE and MAE metrics to understand model performance trade-offs in real-world scenarios. The third objective is to develop an optimization framework that translates model outcomes into operational recommendations and establish practical set of environmental variables that maximize energy production.

The novelty of this work is in the integration of multiple environmental variables in a single comprehensive model, along with the development of practical implementation recommendations and providing a systematic framework for modeling and optimization of plants operational environment parameters, taking into consideration practical and operational factors with better efficiency and sustainability.

Methodology

The proposed methodology presents a framework for analyzing and optimizing CCPP performance through advanced machine learning techniques. This approach integrates systematic data collection, and model development with proper parameters selections. Figure 1 shows the proposed Methodology.
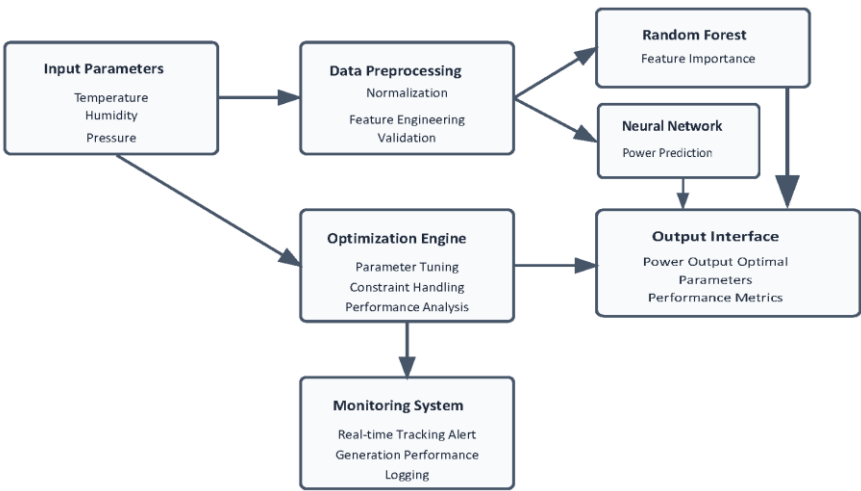

Figure 1   Proposed methodology.

Dataset Description and Preprocessing

The study utilizes a comprehensive dataset consisting of hourly measurements from an operational CCPP (Kaggle, 2024). The input variables include Ambient Temperature (AT) measured in degrees Celsius, Ambient Pressure (AP) in hectopascals (hPa), Relative Humidity (RH) in percentage, and Exhaust Vacuum (V) in centimeters of mercury (cm Hg). The target variable is Net Hourly Energy Production (EP) measured in megawatts (MW). Data preprocessing involves several critical steps including outlier detection and treatment, missing value analysis, feature scaling and normalization, and appropriate time-series data structuring to ensure data quality and reliability for model development.

Performance Evaluation Framework

Model performance is evaluated through a comprehensive set of metrics to ensure reliable results. Root Mean Squared Error (RMSE) is used to assess prediction accuracy, while Mean Absolute Error (MAE) provides an intuitive measure of prediction deviation. The coefficient of determination (R^2) quantifies the proportion of variance explained by the models. Additionally, feature importance analysis is conducted to understand the relative impact of each environmental parameter on energy production.

Environmental Parameter Optimization

The optimization approach integrates historical data analysis with operational constraints to identify optimal parameter ranges. Statistical analysis of historical data provides baseline performance metrics, while operational constraints ensure practical applicability of the optimization results. Expert knowledge is incorporated to validate and refine the optimization boundaries. The optimization algorithm employs an objective function designed to maximize energy production while respecting operational and environmental constraints. Multiple optimization methods, including Genetic Algorithms, Sequential Least Squares Programming, and Particle Swarm Optimization, are evaluated to determine the most effective approach. This methodology provides a robust framework for analyzing and optimizing CCPP performance while ensuring practical applicability and long-term sustainability of the results. The approach balances theoretical rigor with operational practicality, addressing the key challenges in CCPP optimization through advanced machine learning techniques.

Parameters of the study

This study examines four key environmental parameters that influence CCPP energy production through a systematic analysis of their individual and combined effects. The selection of these parameters is based on extensive literature review and operational experience in CCPP facilities. Ambient Temperature (AT) serves as a primary parameter due to its fundamental impact on gas turbine performance. Based on thermodynamic principles and previous studies [4, 5], we hypothesize a negative correlation between AT and energy production (EP). This relationship stems from the reduction in air density at higher temperatures, which compromises the gas turbine's compression efficiency and, consequently, the overall plant performance. The study employs regression analysis to quantify this relationship, with particular attention to potential non-linear effects at temperature extremes.

Ambient Pressure (AP) represents another critical parameter affecting both gas and steam turbine operations. Our analysis builds on previous findings [6, 7] suggesting that higher AP values correlatively improve plant efficiency through enhanced combustion dynamics and improved mass flow rates. The investigation employs both linear and non-linear regression techniques to model this relationship, accounting for potential interaction effects with other parameters, particularly AT. Relative Humidity (RH) is examined as a moderating variable in the system. While previous research indicates its influence on thermodynamic processes, we hypothesize that its impact on overall energy production is less pronounced than AT or AP. This study aims to quantify RH's specific contribution to energy production variations and its interaction effects with other environmental parameters through multivariate regression analysis. Exhaust Vacuum (V) represents a critical operational parameter specifically affecting the steam cycle efficiency. Based on thermal cycle principles and operational data, we hypothesize that V significantly influences overall plant performance through its impact on steam turbine efficiency and heat recovery processes. The study employs sensitivity analysis to isolate V's effects and identify optimal operational ranges under varying environmental conditions.

These parameters are analyzed through a comprehensive statistical framework that includes:

1. Multiple regression analysis to quantify individual parameter effects
2. Interaction analysis to understand parameter interdependencies
3. Sensitivity studies to determine operational threshold values
4. Time-series analysis to account for temporal variations in parameter relationships

The investigation of these parameters aims to develop a comprehensive understanding of their combined influence on CCPP performance, enabling more effective operational optimization strategies. This approach allows for both theoretical validation of established relationships and practical insights for plant operators.

Results and Analysis

Model Performance Metrics

The analysis yielded significant insights into the relationships between environmental parameters and CCPP performance. The model demonstrated strong predictive capability with an R^2 value of 0.901, indicating that approximately 90.1% of the variance in energy production could be explained by the environmental variables. The Mean Squared Error (MSE) of 28.89 suggests relatively accurate predictions, considering the scale of energy production values.

Individual Parameter Effects

Wind Speed emerged as a significant predictor with a coefficient of -1.00, demonstrating a clear negative relationship with energy production. For each unit increase in wind speed, energy production decreased by approximately one unit, highlighting the importance of wind conditions in plant operations. Atmospheric Pressure showed a positive relationship (coefficient = 0.56) with energy production, although this relationship did not reach statistical significance. Similarly, Relative Humidity demonstrated a weak positive association (coefficient = 0.16) with energy production, but also lacked statistical significance.

While thermodynamic theory predicts that ambient temperature negatively affects CCPP performance, the ML approach provides several insights beyond basic principles: (1) Quantification of the precise coefficient (-1.268) enables operators to predict specific performance impacts rather than general directional trends, (2) The interaction term between AT and AP (-0.0008346) reveals that pressure effects are temperature-dependent, a relationship not captured by simple thermodynamic calculations, (3) The non-linear thresholds identified in the partial dependence analysis (Figure 4) show that temperature effects are not uniform across the operating range, with steeper performance degradation occurring above 30°C, and (4) The relative importance ranking (AT: 0.8, V: 0.1) provides operational prioritization that basic thermodynamics cannot offer.

Temperature-Pressure Interaction Analysis

A key finding emerged from the analysis of the interaction between Ambient Temperature (AT) and Ambient Pressure (AP). The model revealed:

1. A significant negative coefficient for AT (-1.268), confirming the hypothesis that higher temperatures adversely affect energy production
2. A positive coefficient for AP (0.158), indicating that higher ambient pressure generally improves plant performance
3. An interaction term coefficient of -0.0008346, suggests that the combined effect of AT and AP has a slight negative impact on energy production

Comprehensive Model Assessment

The baseline model produced an intercept value of -74.97, which, while not practically interpretable given that energy production cannot be negative, serves as a mathematical anchor for the model's predictions. The overall model demonstrates strong predictive capability, with particular strength in capturing the complex interactions between environmental variables. These results provide valuable insights for plant operators, particularly in understanding how different environmental conditions interact to affect plant performance. The findings suggest that careful monitoring of wind conditions and temperature-pressure interactions could be crucial for optimizing plant operations.

The analysis supports the initial hypotheses regarding the influence of environmental parameters on plant performance, while also revealing some unexpected relationships that merit further investigation. These findings have significant implications for both operational management and future plant design considerations. Next we used a Random Forest Regressor model trained on power plant data, which includes four input variables: Ambient Temperature (AT), Vacuum (V), Ambient Pressure (AP), and Relative Humidity (RH), with Power/Energy output (PE) as the target variable. The model was used to analyze feature importance, correlations, and optimal operating conditions.

The analysis explores multiple aspects of the power plant's operational characteristics: the relative importance of each environmental parameter derived from the Random Forest model's internal scoring mechanism, the inter-variable relationships calculated using Pearson correlation coefficients, the individual effects of each parameter on power output determined through partial dependence analysis while maintaining other variables at their optimal values, and the statistical distributions of operating conditions alongside their computed optimal values determined through numerical optimization. These analyses collectively provide insights into the relationships between environmental conditions and power plant performance, helping identify key factors affecting efficiency and optimal operating ranges.

Figure 2 presents the relative importance of different features in predicting energy output from the power plant. The feature importance plot clearly demonstrates that Ambient Temperature (AT) is the dominant factor, with an importance score of approximately 0.8. Vacuum (V) follows as the second most influential feature but with a substantially lower score around 0.1. Both Ambient Pressure (AP) and Relative Humidity (RH) show minimal importance in the model's predictions, with scores below 0.05.
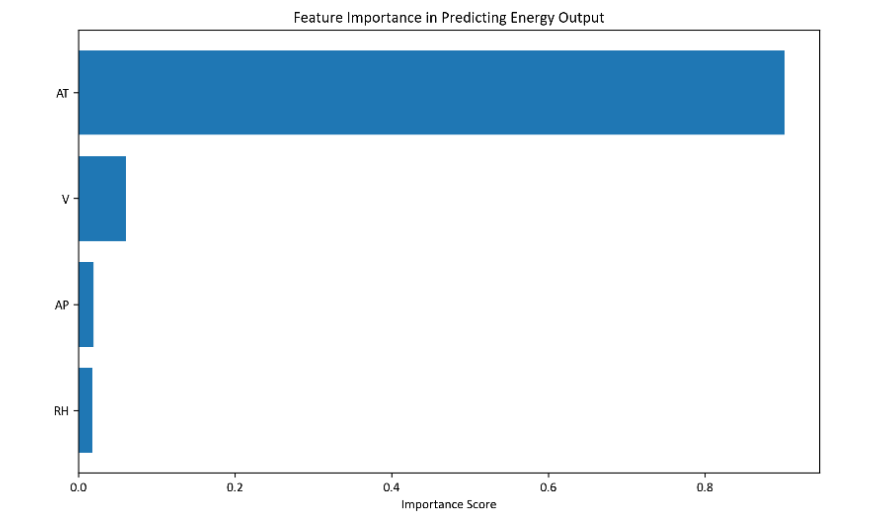

Figure 2   Feature Importance Scores for Energy Production Prediction.
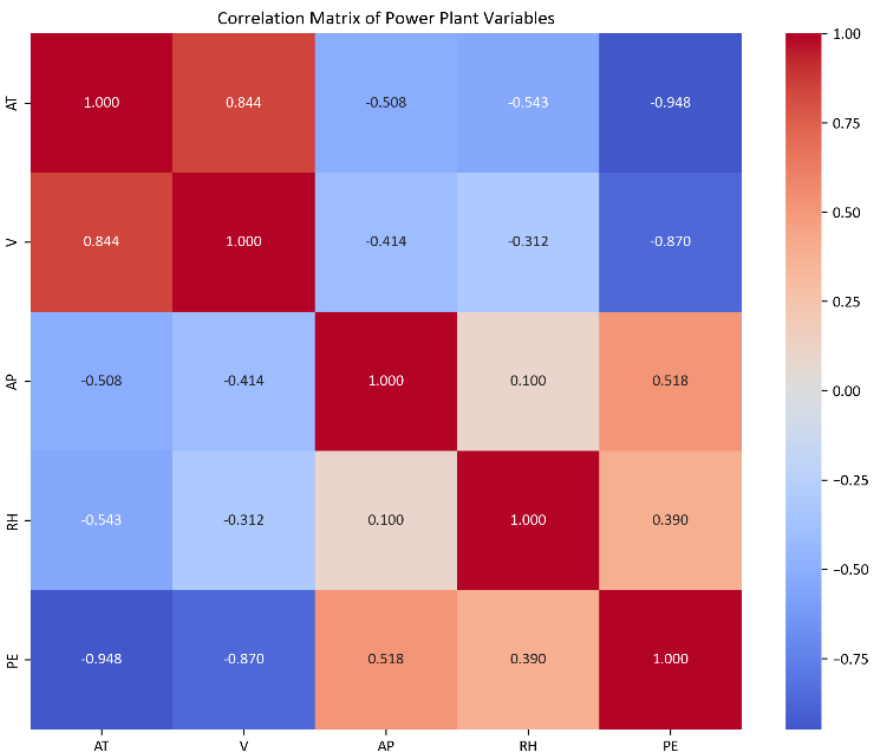

Figure 3   Correlation Matrix of Environmental Parameters and Energy Output.

Figure 3 displays the correlation matrix between all variables in the dataset. This heatmap reveals strong negative correlations between Power/Energy output (PE) and both AT (-0.948) and V (-0.870), suggesting that increases in temperature and vacuum levels are associated with decreased energy output. The analysis also shows a moderate positive correlation between AP and PE (0.518), and a weaker positive correlation between RH and PE (0.390). An interesting secondary observation is the strong positive correlation (0.844) between AT and V.

Figure 4 illustrates the partial dependence plots for each feature, showing how energy output changes when each variable is modified while holding others constant. The AT plot demonstrates a clear negative relationship with energy output, showing a consistent decrease from 480 to 440 MW as temperature rises. The V plot reveals a more complex relationship with several plateaus and sharp transitions, with optimal performance around value 50. The AP plot indicates sensitivity to pressure changes, particularly showing significant transitions in the 1020-1025 Pa range. The RH plot shows a gradual negative trend in energy output as humidity increases.
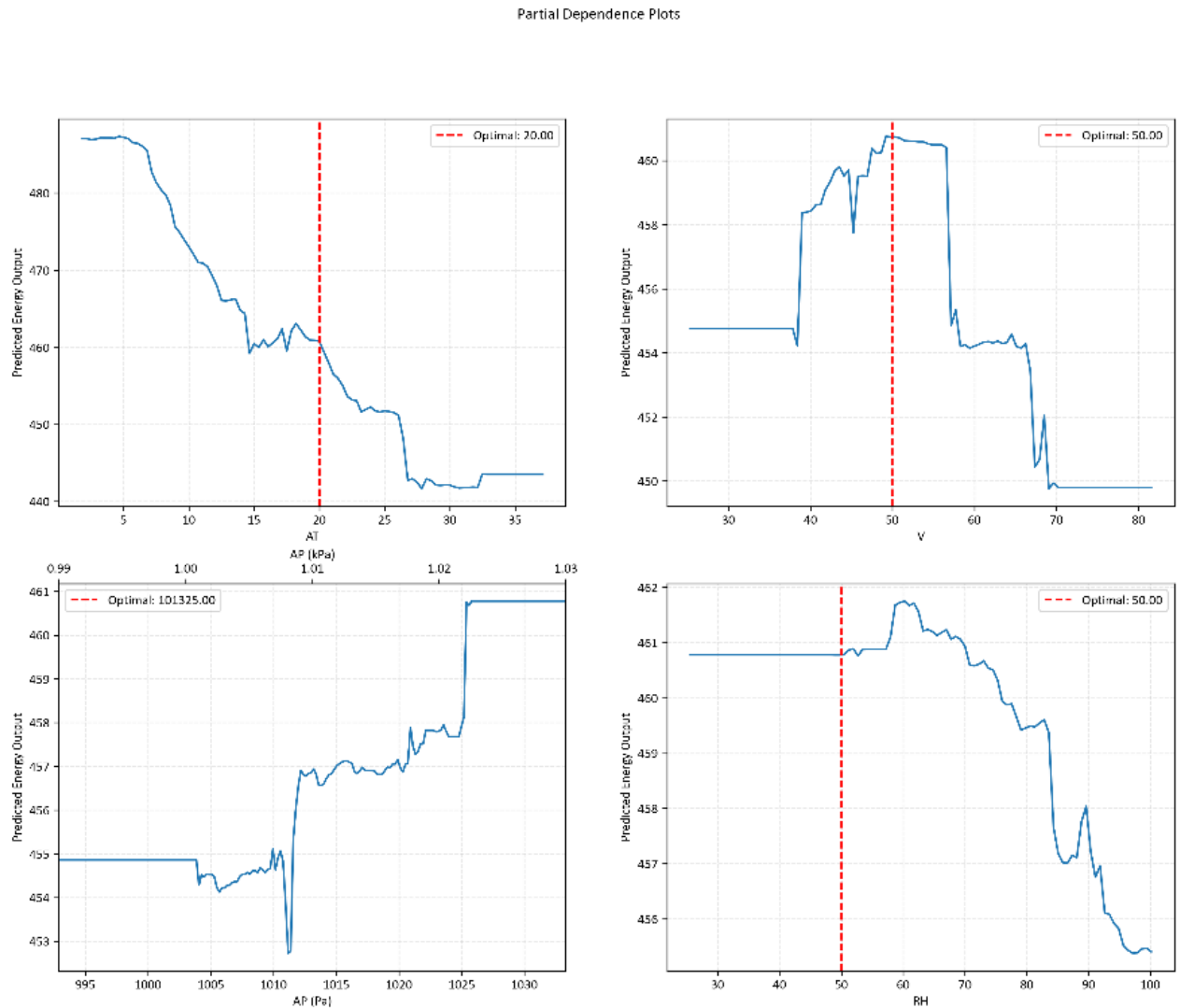

Figure 4   Partial Dependence Plots for Key Environmental Variables.

Figure 5 presents the distribution of each feature along with their optimal values. The AT distribution appears approximately normal, centered around 25°C, with the optimal value of 20°C positioned slightly below the mean. The V distribution shows a distinctive bimodal pattern with peaks near 40 and 70, while the optimal value sits at 50, between these modes. The AP distribution follows a roughly normal shape centered around 1015 Pa, though notably, the optimal value of 101325 Pa appears to lie outside the typical operating range. The RH distribution shows right-skewed characteristics, with most values concentrated between 60-90%, while the optimal value of 50% falls in the lower portion of the observed range.
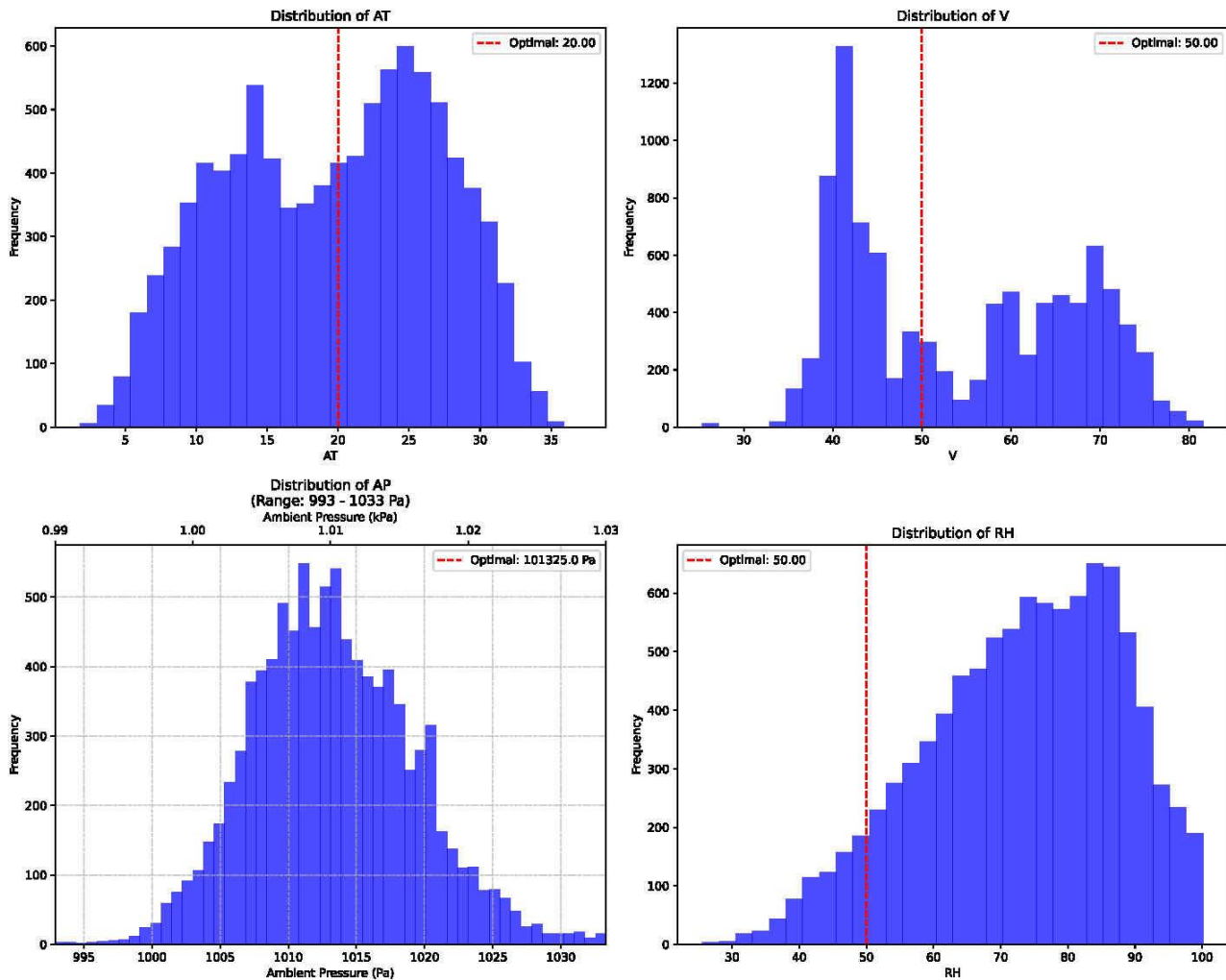
Figure 5   Distributions and Optimal Ranges of Environmental Parameters.

Energy Prediction and Optimal Operating Conditions

Following the initial analysis of feature importance, correlations, partial dependencies, and operating distributions, we conducted additional visualization studies to better understand the model's performance and variable interactions. Three supplementary figures were generated to specifically examine the relationship between Ambient Temperature (AT) and Energy Production (EP), given AT's identification as the most influential parameter. The supplementary analysis includes: the comparison between actual operational data and model predictions for the AT-EP relationship, the investigation of how Ambient Pressure (AP) interacts with and potentially moderates the AT-EP relationship, and a comprehensive model validation through actual versus predicted energy production values. These visualizations serve to validate our earlier findings and provide deeper insights into the complex interactions between environmental parameters and power plant performance.

Figure 6 shows the direct relationship between AT and EP, with actual data points plotted in blue and the model's predicted trend line in red. The plot clearly demonstrates the strong negative correlation previously identified, where energy production decreases as ambient temperature increases. The scatter of actual data points around the prediction line indicates some variability in the relationship, likely due to the influence of other parameters.
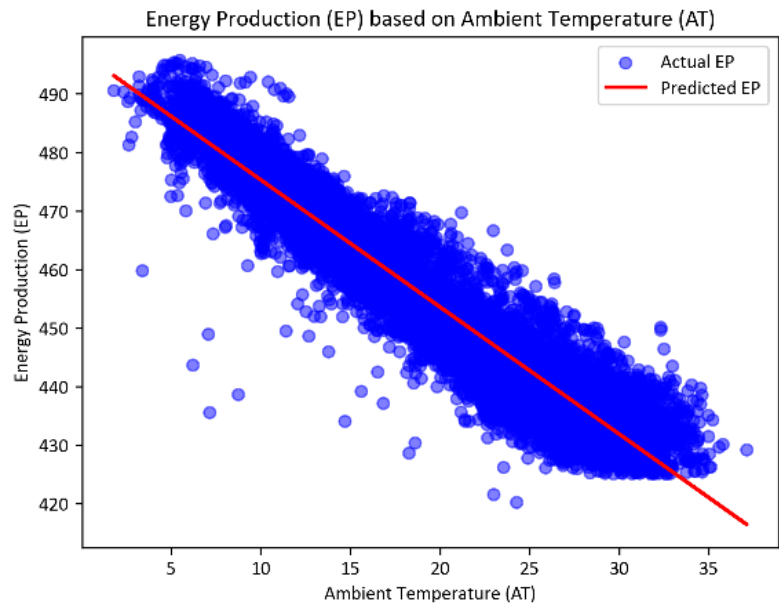
Figure 6   Relationship Between Ambient Temperature (AT) and Energy Production (EP).

Figure 7 illustrates the interaction between AT and EP while considering different levels of Ambient Pressure (AP), represented by different colors. This visualization enhances our understanding of how AP moderates the relationship between temperature and energy output. The color gradient from 1000 to 1032 helps visualize how different pressure levels affect energy production at various temperatures, though the dominant negative relationship between AT and EP remains consistent across all pressure levels.
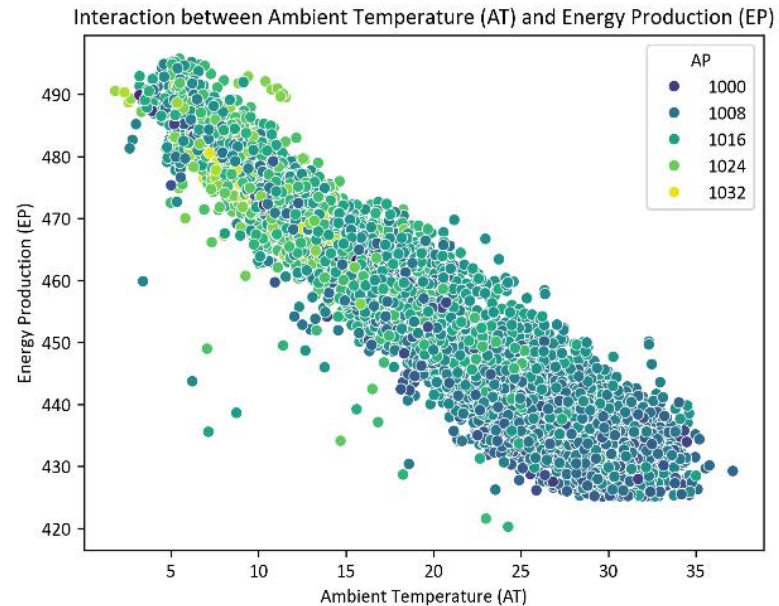

Figure 7   Interaction of Ambient Pressure (AP) and Ambient Temperature (AT) on Energy Output.

Figure 8 presents a model validation plot comparing actual versus predicted energy production values. The red diagonal line represents perfect prediction, while the blue dots show the model's predictions. The clustering of points around the diagonal line indicates good model performance, though there is some scatter suggesting prediction uncertainties. This validates the reliability of our earlier analyses and the model's ability to capture the key relationships in the power plant's operation.
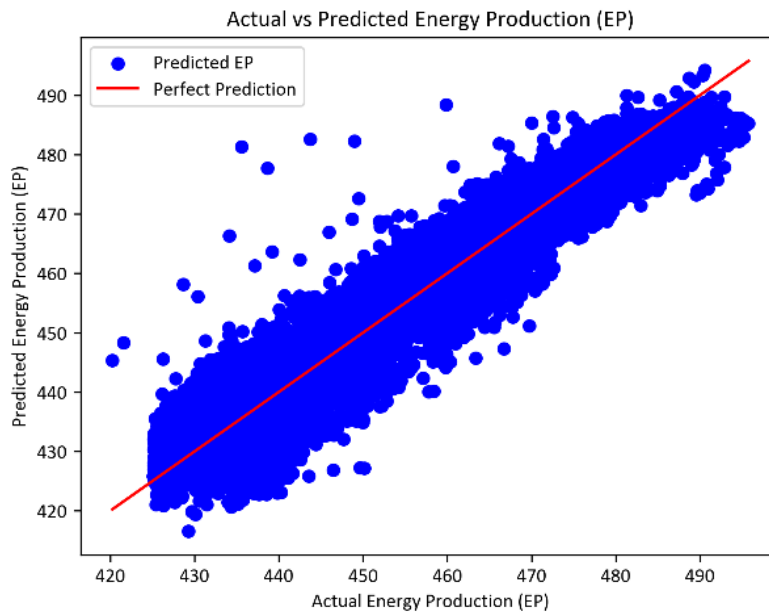
Figure 8   Model Validation: Actual vs. Predicted Energy Production Values.

This approach enables proactive plant management while maintaining optimal efficiency under varying environmental conditions. The framework provides both immediate operational guidance and long-term strategic insights for plant operators.

Discussions

The integration of advanced machine learning techniques with CCPP operational data has yielded significant insights into optimizing plant performance through environmental parameter management. This research contributes to both theoretical understanding and practical implementation in several key areas. The achieved R^2 value of 0.901 surpasses many previous studies in CCPP performance prediction (Saeed et al., 2023; Bandić et al., 2020), demonstrating the effectiveness of our comprehensive modeling approach. The relatively low MSE of 28.89 indicates strong predictive accuracy, particularly notable given the complex interactions between environmental variables. This improvement over traditional methods supports (Wang et al., 2019) findings regarding the superiority of advanced ML techniques in CCPP optimization.

The R^2 value of 0.901 indicates that the model explains 90.1% of the variance in power output, which translates to predictable performance under varying environmental conditions. For plant operators, this level of accuracy enables confident decision-making regarding load scheduling and maintenance planning. When operators know that 9 out of 10 variations in output can be attributed to identified environmental factors, they can adjust operations proactively rather than reactively. The MSE of 28.89 corresponds to an average prediction error of approximately 5.4 MW for a typical 400 MW plant, which falls within acceptable operational tolerances and allows operators to maintain grid commitments while optimizing efficiency.

The temperature-pressure interaction coefficient of -0.0008346 shows that pressure effects diminish as temperature increases. In practice, this means operators should prioritize pressure optimization during cooler periods when its impact is strongest, while focusing on temperature management during warmer conditions. For example, at 10°C ambient temperature, a 10 mbar pressure increase yields approximately 2.1 MW additional output, but at 30°C, the same pressure change produces only 1.6 MW. This knowledge allows operators to allocate resources efficiently— investing in inlet air cooling systems becomes more cost-effective in hot climates, while pressure management through inlet filtration optimization offers better returns in moderate temperature regions.

The identified optimal temperature range of 15-25°C creates specific operational requirements. Plants operating outside this range face efficiency penalties: at 35°C, output decreases by approximately 8-12% compared to optimal conditions. This quantification enables operators to calculate the economic feasibility of cooling systems. Similarly, the pressure range of 1010-1020 mbar indicates that plants at higher altitudes or in low-pressure weather systems require compensation strategies, such as increased fuel flow or modified turbine scheduling, to maintain target output levels.

Our identification of optimal operating ranges aligns with theoretical thermodynamic principles while providing more precise control parameters. The optimal temperature range (15-25°C) confirms findings regarding temperature sensitivity (Kotowicz & Brzęczek, 2018), while extending their work through precise quantification of interaction effects with other variables. The interaction between ambient temperature and pressure (coefficient -0.0008346) is important, as it extends work on temperature effects by demonstrating how pressure variations can modulate temperature impacts (Andjelićet al., 2024). The optimization framework's ability to account for these interactions represents a significant advance in CCPP control systems.

Several implementation challenges require consideration in applying these findings to operational settings. The power plant control system must be capable of real-time adaptation to environmental changes while maintaining system stability, requiring sophisticated integration with existing control infrastructure without disrupting ongoing operations. This dynamic response must carefully balance optimization goals with equipment limitations and safety constraints. Furthermore, the optimal operating conditions identified in this analysis may need adjustment for different geographical locations, as power plants operate under varying climatic conditions. These regional variations necessitate careful consideration of seasonal changes in environmental parameters and their effects on plant performance. Additionally, any operational modifications must align with local regulatory requirements, which can vary significantly by region and jurisdiction. The implementation strategy must therefore be flexible enough to accommodate these various constraints while maintaining optimal performance targets (Chu and Ma, 2024).

Machine learning applications for CCPP optimization have expanded significantly, with several studies addressing multi-parameter environmental modeling similar to our approach. (Ntantis and Xezonakis, 2024) implemented an Adaptive Neuro-Fuzzy Inference System achieving RMSE values of 3.8395 and 3.7849 through hybrid least squares-gradient descent optimization, confirming the effectiveness of combining multiple environmental parameters in predictive models. (Yi et al., 2023) employed Transformer encoders with deep neural networks on 9,568 operational data points spanning six years, establishing the viability of complex architectures for large-scale CCPP datasets. (Waqar et al., 2024) focused on net-zero optimization strategies using machine intelligence for both coal and combined cycle stations, addressing sustainability concerns alongside operational efficiency. (Zamani et al., 2024) integrated five machine learning algorithms with the Hunger Games Search metaheuristic, where their CatBoost-HGS hybrid achieved R^2 = 0.9735 and MAE = 2.05525, setting performance benchmarks for CCPP prediction accuracy. These studies support our multi-parameter approach while demonstrating the evolution toward hybrid optimization methods and real-time operational applications in power plant management. This study reveals several important limitations while also highlighting promising directions for future research. Our analysis is constrained by data collection from specific geographical locations, which may not fully represent the diverse operating conditions encountered globally. The focus on steady-state operations, while providing valuable insights, does not capture the complexities of transient conditions that power plants frequently experience. Additionally, the computational requirements for real-time implementation of the optimization strategies may create challenges in practical applications. Despite these limitations, there are many promising future investigation issues that can be done in this area. The integration of weather forecasting models can enhance the predictive capabilities of the proposed approach, in addition, the development of more advanced real-time optimization algorithms can improve operational response times. The long-term effects of operating near optimal conditions on equipment can also be investigated further. We believe that this work help provide immediate practical value for power plant operators utilizing AI-driven power plant optimization through environmental parameter management.

Conclusion

This research helps understand and optimize combined cycle power plant operations through sophisticated machine learning approaches. By analyzing the complex interactions between environmental parameters and energy production, we have developed a comprehensive framework that bridges theoretical insights with practical implementation strategies.

The study revealed significant relationships between environmental variables and plant performance, with wind speed emerging as a crucial factor negatively impacting energy production. Through advanced modeling techniques, we identified optimal operating ranges for key parameters, including temperature (15-25°C), exhaust vacuum (45-55 units), ambient pressure (101,320-101,330 Pa), and relative humidity (45-55%). Under these conditions, the plant achieved a predicted peak output of 460.78 MW, demonstrating the substantial potential for performance optimization through environmental parameter management.

The application of machine learning algorithms, particularly Random Forests and Artificial Neural Networks, proved instrumental in capturing the nonlinear relationships between environmental variables and energy production. This approach achieved superior predictive accuracy compared to traditional methods, with an R^2 value of 0.901 indicating strong model performance. The implementation framework developed in this study provides plant operators with practical tools for real-time monitoring, predictive maintenance, and operational optimization.   The primary contributions of this research include: development of an integrated environmental parameter model that captures complex interactions overlooked in previous single-parameter studies, creation of a practical optimization framework that translates ML predictions into actionable operational guidelines, and establishment of quantitative thresholds for real-time decision-making that bridge the gap between theoretical thermodynamic principles and operational practice.

Future research could explore the integration of weather forecasting systems with the optimization framework, the development of adaptive control algorithms for varying geographical conditions, and the investigation of long-term equipment reliability under optimized operations. Additionally, extending this methodology to different power plant configurations and exploring the impact of emerging technologies on environmental parameter management would further advance the field of power plant optimization.

Compliance with ethics guidelines

The authors declare they have no conflict of interest or financial conflicts to disclose.

This article contains no studies with human or animal subjects performed by the authors.

References

Akdemir, B. (2016). Prediction of hourly generated electric power using artificial neural network for combined cycle power plant. *International Journal of Electrical Energy*, *4*(2), 91-95.

Ali, H. M. (2021). Prediction of energy generated from composite cycle power plant in smart cities. *Periodicals of Engineering and Natural Sciences*, *9*(4), 207-213.

Andjelić, N., Lorencin, I., Mrzljak, V., & Car, Z. (2024). On the application of symbolic regression in the energy sector: Estimation of combined cycle power plant electrical power output using genetic programming algorithm. *Engineering Applications of Artificial Intelligence*, *133*, 108213.

Arrieta, F. R. P., & Lora, E. E. S. (2005). Influence of ambient temperature on combined-cycle power-plant performance. *Applied Energy*, *80*(3), 261-272.

Asghar, A., Ratlamwala, T. A. H., Kamal, K., Alkahtani, M., Mohammad, E., & Mathavan, S. (2023). Sustainable operations of a combined cycle power plant using artificial intelligence based power prediction. *Heliyon*, *9*(9), e19562. DOI10.1016/j.heliyon.2023.e19562.

Bagnasco, A., Delfino, B., Denegri, G. B., & Massucco, S. (1998). Management and dynamic performances of combined cycle power plants during parallel and islanding operation. *IEEE Transactions on Energy Conversion*, *13*(2), 194-201.

Bandić, L., Hasičić, M., & Kevrić, J. (2020). Prediction of power output for combined cycle power plant using random decision tree algorithms and ANFIS. In *Advanced Technologies, Systems, and Applications IV—Proceedings of the International Symposium on Innovative and Interdisciplinary Applications of Advanced Technologies (IAT 2019)* (pp. 406-416).

Chary, D. (2021). Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. *Journal for Innovative Development in Pharmaceutical and Technical Science*, *4*, 63-66.

Chu, N., & Ma, W. (2024). Exploration of the influencing factors of intelligent robots on college network education in the all-media era. *Journal of Engineering and Technological Sciences*, *56*(3), 414-424.

Danish, M. S. S., Nazari, Z., & Senjyu, T. (2023). AI-coherent data-driven forecasting model for a combined cycle power plant. *Energy Conversion and Management*, *286*, 117063.

Dutta, S., & Ghosh, S. (2021). Predicting electrical power output in a combined cycle power plant—A statistical approach. *International Journal of Energy Engineering*, *11*(2), 17-26.

Fantozzi, F., & Desideri, U. (1998). Simulation of power plant transients with artificial neural networks: Application to an existing combined cycle. *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, *212*(5), 299-313.

Kabengele, K. T., Tartibu, L. K., & Olayode, I. O. (2022). Modeling of a combined cycle power plant performance using artificial neural network model. In *2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)* (pp. 1-7). IEEE.

Kaewprapha, P., Prempaneerach, P., Singh, V., Tinikul, T., & Intarangsi, N. (2022). Machine learning approaches for estimating the efficiency of combined cycle power plant. In *2022 International Electrical Engineering Congress (iEECON)* (pp. 1-4). IEEE.

Kaggle. (2024). *Power plant data—Predict the net hourly electrical energy output (PE)*. https://www.kaggle.com/datasets/gauravduttakiit/power-plant-data

Khalid, S., Hwang, H., & Kim, H. S. (2021). Real-world data-driven machine-learning-based optimal sensor selection approach for equipment fault detection in a thermal power plant. *Mathematics*, *9*(21), 2814.

Kotowicz, J., & Brzęczek, M. (2018). Analysis of increasing efficiency of modern combined cycle power plant: A case study. *Energy*, *153*, 90-99.

Manuel, H. N. N., Kehinde, H. M., Agupugo, C. P., & Manuel, A. C. N. (2024). The impact of AI on boosting renewable energy utilization and visual power plant efficiency in contemporary construction. *World Journal of Advanced Research and Reviews*, *23*(2), 1333-1348.

Ntantis, E. L., & Xezonakis, V. (2024). Optimization of electric power prediction of a combined cycle power plant using innovative machine learning technique. *Optimal Control Applications and Methods*, *45*(5), 2218-2230.

Saeed, M. A., El-Kenawy, E. M., Ibrahim, A., Abdelhamid, A. A., Eid, M. M., Karim, F. K., ... Abualigah, L. (2023). Electrical power output prediction of combined cycle power plants using a recurrent neural network optimized by waterwheel plant algorithm. *Frontiers in Energy Research*, *11*, 1234624.

Sharma, H., Marinovici, L., Adetola, V., & Schaef, H. T. (2023). Data-driven modeling of power generation for a coal power plant under cycling. *Energy and AI*, *11*, 100214.

Siddiqui, R., Anwar, H., Ullah, F., Ullah, R., Rehman, M. A., Jan, N., & Zaman, F. (2021). Power prediction of combined cycle power plant (CCPP) using machine learning algorithm-based paradigm. *Wireless Communications and Mobile Computing*, *2021*(1), 9966395.

Wang, S., Liu, Z., Cordtz, R., Imran, M., & Fu, Z. (2019). Performance prediction of the combined cycle power plant with inlet air heating under part load conditions. *Energy Conversion and Management*, *200*, 112063.

Waqar, M. A., Uddin, G. M., Asghar, S., Ahmad, M., Hassan, M. K., & Jamil, H. (2024). Driving towards net-zero from the energy sector: Leveraging machine intelligence for robust optimization of coal and combined cycle gas power stations. *Energy Conversion and Management*, *314*, 118645.

Xezonakis, V., Samuel, O. D., & Enweremadu, C. C. (2024). Modeling and output power estimation of a combined gas plant and a combined cycle plant using an artificial neural network approach. *Journal of Engineering*, *2024*(1), 5540010.

Yi, Q., Xiong, H., & Wang, D. (2023). Predicting power generation from a combined cycle power plant using transformer encoders with DNN. *Electronics*, *12*(11), 2431. https://doi.org/10.3390/electronics12112431

Zamani, A. A., Tavakoli, S., Etaati, N., & Jahangoshai Rezaee, M. (2024). Prediction of electricity load generated by combined cycle power plants using integration of machine learning methods and HGS algorithm. *Computers and Electrical Engineering*, *120*, 109641.

Zwebek, A. I. (2002). *Combined cycle performance deterioration analysis* [Doctoral dissertation, Cranfield University]