

Article

Prediction Modeling of Flue Gas Control for Combustion Efficiency Optimization for Steel Mill Power Plant Boilers Based on Partial Least Squares Regression (PLSR)

Sang-Mok Lee ^{1,2}, So-Won Choi ¹ and Eul-Bum Lee ^{1,3,*}

¹ Graduate Institute of Ferrous and Eco Materials Technology, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea; lopes004@postech.ac.kr (S.-M.L.); smilesowon@postech.ac.kr (S.-W.C.)

² Energy Technology Section, Energy Department, Pohang Iron and Steel Company (POSCO), Pohang 37754, Republic of Korea

³ Department of Industrial and Management Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea

* Correspondence: dreblee@postech.ac.kr; Tel.: +82-54-279-0136

Abstract: The energy-intensive steel industry, which consumes substantial amounts of electricity, meets its power demands through external electricity purchases and self-generation through the operation of its own generators. This study aimed to optimize boiler combustion efficiency and increase power generation output by deriving optimal operational values for O₂ and CO within the boiler flue gas using machine learning (ML) with the aim of achieving maximum boiler efficiency. This study focuses on the power-generation boilers at steel mill P in Korea. First, 361 types of operation data from power generation equipment were collected and preprocessed. Subsequently, a partial least squares regression (PLSR) algorithm was used to develop a prediction model for O₂ and CO values, known as the Boiler Flue Gas Prediction Model (BFG-PM). The prediction accuracy for O₂ was notably high (83.2%), whereas that for CO was lower (53.4%). Nonetheless, the model's reliability was high because more than 90% of the predicted values were within a 10% error range. Finally, the correlation of the BFG-PM model was applied to the performance test code (PTC) 4.0 for the boiler efficiency calculations formula, deriving the optimal O₂ and CO control points. Through a simulation, it was verified that the boiler efficiency was improved by controlling the combustion air. In addition, an average increase in boiler efficiency of 0.29% was confirmed by applying it directly to the generator operating on-site. The results of this study are expected to contribute to annual cost savings, with a reduction of USD 217,000 in electricity purchasing costs and USD 19,700 in greenhouse gas emissions trading expenses.

Keywords: machine learning; power plant in steel mill; boiler efficiency; combustion control; flue gas prediction; regression; partial least squares; performance test code 4.0

Citation: Lee, S.-M.; Choi, S.-W.; Lee, E.-B. Prediction Modeling of Flue Gas Control for Combustion Efficiency Optimization for Steel Mill Power Plant Boilers Based on Partial Least Squares Regression (PLSR). *Energies* **2023**, *16*, 6907. <https://doi.org/10.3390/en16196907>

Academic Editors: Silvia Maria Zanolli, Crescenzo Pepe

Received: 3 September 2023

Revised: 25 September 2023

Accepted: 27 September 2023

Published: 30 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background of Study

An increase in energy demand and energy costs since the COVID-19 crisis has affected the production cost of power generation, in which the unit cost of electricity power in Korea has increased by 64.4%, six times since January 2021 [1]. This results in a burden of production costs in the manufacturing industry, particularly in the steel-making industry, which consumes a great deal of energy. Most of the electric power used in a representative steel mill, P, is supplied by in-house power stations; however, an insufficient amount of power is purchased from external suppliers. The total annual power consumption of steel mill P was 24,492 GWh, which corresponds to the amount of power generated

by three nuclear reactors. Specifically, 16,013 GWh (65%) is generated by byproduct gas, 2979 GWh (12%) is generated by LNG, 2653 GWh is generated by other generation sources, and a deficient amount of 2847 GWh is procured through external purchases [2]. Therefore, Steel Mill P needs an increased power generation output because it still purchases power from external sources despite its own power generation operation.

Power generation efficiency is mainly determined by the boiler and turbine efficiency. A steam power-type boiler generates steam by burning fuels at 90% thermal efficiency, whereas a steam turbine receiving the generated steam produces 40% power generation efficiency by converting high-temperature, high-pressure force into electricity at 50% thermal efficiency [3]. In the past, process improvements were achieved through exhaust heat recovery or by replacing the balance of the plant (BOP) with high-efficiency equipment.

Generators operated at steel mill P are divided into two types: steam power generation, in which steam turbines are operated through high-temperature, high-pressure steam using by-product gas; and combined power generation, in which gas turbines are operated using liquefied natural gas (LNG), and additional power is generated through the exhaust gas heat of the gas turbines [4]. Of these two types, this study targeted steam power generation, for which boiler combustion control is difficult because changes in calories are large owing to various fuel types and the supplied flow rate varies frequently. Steam power generation utilizes four types of byproduct gases: blast furnace gas (BFG), coke oven gas (COG), Linz–Donawitz converter gas (LDG), and FINEX off-gas (FOG). These gases are generated during the iron-making, steel-making, and formation processes. Table 1 presents the calories and components of each type of byproduct gas.

Table 1. By-product gas calorie content and components.

Category		BFG ¹	COG ²	LDG ³	FOG ⁴
Standard calorie (Kcal/Nm ³)		750	4400	2000	1350
Specific gravity		1.03	0.42	1.04	1.38
Component (%)	CO ₂	20.7	3.1	17.8	33
	O ₂	-	0.3	-	-
	C ₂ H ₄	-	2.0	-	-
	CO	20.0	8.4	64.2	43
	CH ₄	-	26.6	-	1
	H ₂	3.2	56.4	2.0	21
	N ₂	54.1	2.3	15.9	2

¹ BFG: blast furnace gas. ² COG: coke oven gas. ³ LDG: Linz–Donawitz converter gas. ⁴ FOG: FINEX off-gas.

The power generation operation process using the byproduct gas is as follows. First, fuel and appropriate combustion air are burned in a boiler, and then the tube positioned in the top part of the boiler is heated to convert water into high-temperature, high-pressure steam of 541 °C and 131 bar. Here, the high-temperature gas burned in the boiler is discharged externally through a duct and chimney, and a gas–air heater (GAH) for high-temperature heat recovery is installed to heat the combustion air. High-temperature, high-pressure steam moves to a steam turbine through pipes to rotate the turbine wings, and the generator connected to the turbine form a magnetic field by inducing an interaction between the rotor and stator through rotation, thus generating electricity. Figure 1 shows the power generation process, which consists of three parts: a boiler that burns the by-product gas, a generator that produces electric power, and the BOP.

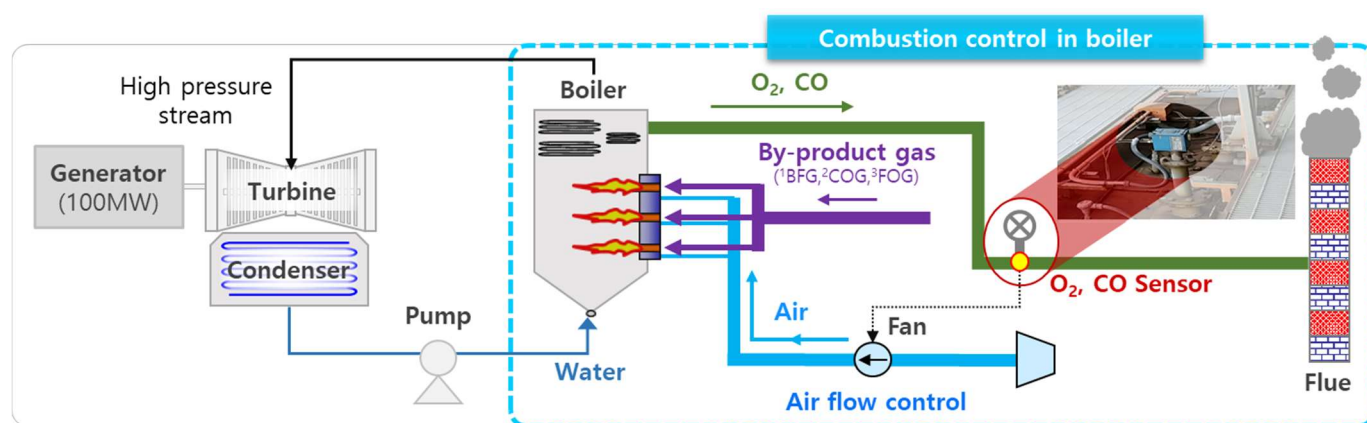


Figure 1. Power generation process in a steel mill (¹ BFG: Blast furnace gas, ² COG: Coke oven gas, ³ FOG: FINEX off-gas).

The boiler operation can become highly efficient by optimizing combustion control, which is adjustable based on the air–fuel ratio (AFR), as shown in combustion theory [5]. Boiler efficiency can be verified in real-time if there is a system, or combustion conditions can be monitored and controlled by checking the O₂ and CO amounts in the flue gas discharged through an exhaust air duct if a system is not available. In other words, high-efficiency operation is feasible if combustion is controlled by the amounts of O₂ and CO in the flue gas. Therefore, a boiler can be operated with optimal efficiency through an ML analysis performed using the boiler combustion efficiency, AFR, and the O₂ and CO amounts in the flue gas.

1.2. Problem Statement and Objectives

The steam-powered boiler at steel mill P involves combustion control based on the following mechanism for high-efficiency operation but also entails several problems. The combustion control mechanism is as follows: when fuel and air are placed in a burner together and ignited, combustion gas is generated. The O₂ and CO of the combustion gas are analyzed to determine if combustion has been properly carried out; if the amount of O₂ is lower or that of CO₂ is higher than the threshold, more combustion air is added because it is determined as incomplete combustion when the fuel has not been sufficiently burned. Contrarily, if the amount of CO₂ in the combustion air is excessive, combustion air input is adjusted since it is determined that heat loss has occurred due to excessive air input. Figure 2 shows the combustion control mechanism of a power plant boiler.

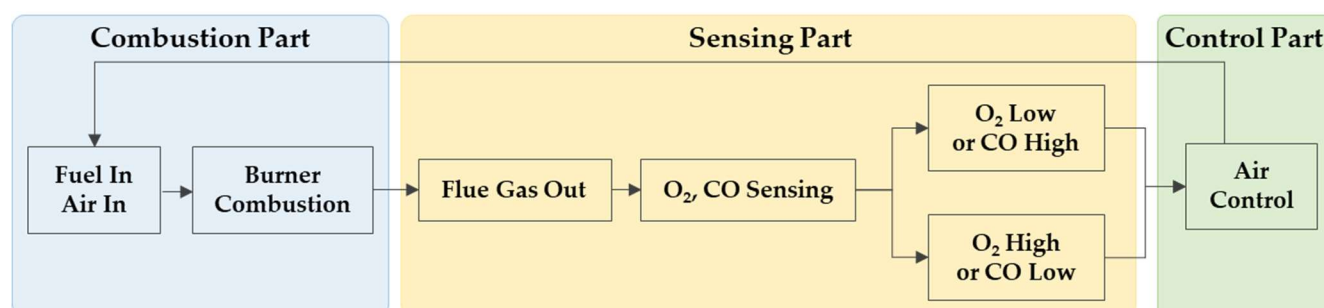


Figure 2. Combustion control mechanism of power plant boiler.

Currently, power plant boilers have issues with combustion control owing to the following problems:

- Problem 1: Insufficient reliability of O₂ and CO analyzer [6];
- Problem 2: Combustion air is not controlled [7];

- Problem 3: Absence of flue gas control point for optimizing boiler combustion efficiency [8].

First, the insufficient reliability of the O₂ and CO analyzers in the sensing part is caused by the contraction and expansion of the duct as the high-temperature flue gas passes through the exhaust duct. Furthermore, frequent malfunctions and lower reception rates are caused by vibrations generated by the boiler, which affects the analyzer. Unlike other power generation fuels, by-product gases contain a large amount of dust because they are by-products of steel-making processes. Therefore, a blockade in the zirconia analyzer for analyzing a specific point occurred because of the dust contained in the gas, in addition to hunting and peak analysis values.

Second, the combustion air is not controlled in the control part because of the reduced reliability of the analyzer and the inability to verify the amounts of O₂ and CO in the flue gas. An operator ensures that a constant amount of air is placed without controlling the combustion air, even when the operation state changes. In other words, changes in the type or supply of by-product gas require changes in the AFR control as well; however, the initial setting of ensuring that flue gas O₂ becomes 2% is not adjusted, which results in excess air combustion or incomplete combustion, depending on the context. Therefore, uncertain O₂ management causes incomplete combustion of the boiler or excess air combustion, leading to reduced boiler efficiency and power generation output.

Finally, the problem with the absence of a control point for optimizing boiler combustion efficiency is described. Currently, operation with 2% O₂ is set as the standard, based on equipment conditions and previous experience [9]. However, it is impossible to verify whether the boiler efficiency has improved, even when the operation ensures 2% O₂. If a system is available for checking boiler efficiency, combustion air can be controlled by deducing the amounts of O₂ and CO in the flue gas that optimizes the boiler efficiency, even if the fuel flow rate and calories change; however, such a system is currently unavailable. Therefore, there is an absence of a system for examining boiler efficiency.

As a part of the solution, the analyzer was replaced with the latest high-performance sensor. A tunable diode laser spectrometer (TDLS) capable of a more stable analysis across a broader range was additionally installed; securing reliability was still limited because the O₂ concentration varied depending on the analysis point owing to the stratification phenomenon where the flue gas is not easily mixed. In addition, a dust collector using the cyclone principle was installed, and chemicals were sprayed to partially remove the dust included in the by-product gas supplied to the power plant [10]. However, the dust removal effect was insignificant, and other maintenance limitations remained.

To solve the problem of inadequate control of boiler combustion, this study aimed to perform combustion control by securing the reliability of the O₂ and CO analyzers to optimize boiler combustion efficiency and increase the power generation output. For the above purposes, the combustion condition of a boiler needs to be inspected first, based on a stable prediction of the amounts of O₂ and CO, even when the analyzer is not operating smoothly, by developing a model for predicting the amounts of O₂ and CO in the flue gas in a boiler. To this end, ML modeling was designed to find critical variables related to the boiler flue gases, O₂ and CO, and to predict the final values based on long-term micro data collected from various internet of things (IoT) devices installed on the generators at steel mill P. Because a generator is highly complex equipment and the types of data collected from sensors vary markedly, it is crucial to find the major variables related to the predicted values. This study developed a model based on data collected from sensors attached to the power generation facilities. The sensor data are mainly composed of continuous numerical values and exhibit time-series patterns over time. In addition, these data are characterized by their large volume. Hence, considering the importance of selecting key variables for modeling, this study developed a model using a data-driven approach based on machine learning, specifically the partial least squares regression (PLSR), which allows for assessing variable importance.

To guarantee the effectiveness of increased power generation output, this study aimed to deduce the amounts of flue gas O_2 and CO that maximized boiler combustion efficiency for each fuel condition based on a boiler combustion efficiency calculation equation. The intention was to enable combustion air control by applying boiler flue gas O_2 and CO model values derived using ML modeling. Consequently, operators would conduct operations based on highly reliable flue gas conditions, thereby increasing the boiler combustion efficiency and enhancing the power generation output. This approach minimizes the purchase of external electricity by optimizing power generation within the facility.

This study distinguishes itself from previous research in several key ways. The novelty of this study lies in its utilization of machine learning not only to predict flue gas O_2 levels but also CO levels, and additionally in integrating research on optimizing power plant boiler efficiency. This integrated approach provides more accurate information for boiler operation and has the potential to yield superior results in terms of efficiency optimization. Furthermore, this research has systematically applied the developed boiler flue gas prediction model to an actual operational site in a steel mill. This validation of practical utility extends beyond theoretical models, confirming its applicability in real-world environments. Moreover, the analysis of the financial impact of the developed model is another unique aspect of this research, demonstrating the economic benefits of the developed technology.

While various studies have explored the application of machine learning and AI in optimizing combustion efficiency in power plant boilers [11–31], most have been limited to either focusing solely on combustion flue gas O_2 [19–23] or exclusively on boiler combustion efficiency optimization [24–31]. This research sets itself apart by deriving boiler efficiency optimization points using predictions of both combustion flue gas O_2 and CO levels, highlighting a scientific gap in existing research. These characteristics highlight that this research is technologically advanced and holds the potential to make a significant contribution to real-world industrial applications.

1.3. Literature Review

To improve the boiler combustion efficiency and increase the power generation output, which are the objectives of this study, previous studies conducted on traditional methods were reviewed, and flue gas monitoring technology to check the boiler combustion state was examined. Subsequently, cases that integrated artificial intelligence (AI) with the boiler combustion efficiency optimization performed in this study were reviewed, and the necessity of this study was highlighted based on the implications and limitations of previous studies.

1.3.1. Conventional Studies on Improvements in Power Plant Efficiency

Because generators are large-scale general equipment, overall efficiency improvement can be attained by identifying the energy loss and performance degradation per unit of equipment or by rationalizing or replacing equipment with the latest high-efficiency equipment. Because related efforts have been made in the past and are still ongoing, previous studies that deduced improvements in the energy loss and efficiency of generators were reviewed. Murehwa et al. analyzed the loss of available energy, or exergy, of each equipment to identify energy loss for determining improvement measures for plant processes and discovered that the greatest loss of available energy occurred in the boiler (48.92%), mostly as a result of the combustion reaction and a large temperature difference during heat transfer between combustion gas and steam. The researchers also identified tube contamination, burner defects, fuel quality, inefficient exhaust ventilation, and air heater contamination. Therefore, modifying or replacing boilers is recommended to improve power generation efficiency by minimizing energy loss [11].

Regarding the energy loss of a boiler, which is directly related to power generation efficiency, Karri discovered that efficiency degradation was mainly caused by heat loss

due to flue gas loss and moisture content of fuel, and that the performance of a boiler could be improved by decreasing flue gas temperature and lowering the moisture content of fuel through boiler efficiency calculation [12]. Mandi et al. conducted a study to improve the overall efficiency by increasing the energy efficiency of major auxiliary devices of a generator. The impact of a decrease in the induced draft fan (IDF) load after IDF overhaul and flue gas pressure decrease on overall efficiency improvement was determined [13]. Hasanuzzaman et al. proposed a power generation efficiency improvement through energy reduction using a recuperator for preheating the combustion air and an economizer for waste heat recovery [14]. Ibrahim et al. proposed a parametric analysis and simulation model to improve the performance and power output of a power plant and discovered that increased steam pressure induced an increase in the overall thermal efficiency and power output, and that the turbine inlet temperature was a key variable [15]. Errami et al. presented a nonlinear control strategy based on nonlinear backstepping theory for tracking the maximum power point and regulating the rotational speed of the permanent magnet synchronous generator (PMSG) in a 4 MW wind power generation system [16]. Mosobi and Gao established a distributed generator (DG) system for addressing power quality (PQ) issues in low-voltage networks based on renewable energy sources. This DG system integrates components such as a photovoltaic system (PVS), wind energy generating system (WEGS), and micro-hydro generating system (MHGS), and connects them to a common DC bus [17]. Regarding efficiency improvement for other equipment in addition to the boiler, Khaleel et al. verified the improvement in efficiency by modeling using an exergy evaluation to ensure that coal-fired electrical power plants could still be operated optimally under new conditions instead of the initial design conditions, as conditions can deteriorate [18].

1.3.2. Studies on Boiler Flue Gas Prediction

Boiler combustion efficiency optimization for increasing power generation efficiency can be carried out when combustion flue gas can be consistently and accurately measured. Thus, a variety of combustion flue gas monitoring methods have been examined because of the limitations in measuring O₂ and CO amounts in combustion flue gas using a physical analyzer. Zaporozhets developed linear model (LM) programmer-based flue gas O₂ concentration monitoring software to monitor the combustion process of a boiler, as well as other types of software that were applicable to the combustion process control of various types of fuels [19]. Tang et al. extracted highly correlated variables using the LASSO algorithm to predict the amount of O₂ in boiler combustion flue gas, using a deep belief network (DBN) to model two variable groups, and combined forecasting using least-squares support vector machines (LS-SVM) [20]. Pan et al. predicted the O₂ amount in boiler combustion flue gas using long short-term memory (LSTM) modeling within the Keras deep learning framework and further augmented the model's accuracy by selecting parameters through experiments [21]. As a forecasting method using AI for measuring combustion flue gas, Effendy et al. proved that the O₂ amount in combustion flue gas could successfully be predicted by using an artificial neural network (ANN) and a random forest-based soft sensor for monitoring the boiler combustion efficiency [22]. Li et al. suggested a convolutional neural network (CNN)-based model to improve the accuracy of predicting the O₂ amount in boiler combustion flue gas and verified high consistency through comparisons [23].

1.3.3. Boiler Combustion Efficiency Improvement Using AI

This study also reviews previous studies that utilized AI to optimize the combustion efficiency of power-generation boilers. Santoso et al. proposed an approach for designing a fuzzy logic controller for the optimization of the air–fuel ratio during the boiler combustion process and suggested boiler combustion efficiency improvement through the reduction of excessive air [24]. Liu et al. proposed a method for improving the boiler combustion efficiency by integrating a non-dominated sorting genetic algorithm (NSGA2) of multi-

objective optimization techniques with computational fluid dynamics (CFD) because conventional AI-based optimization techniques were limited, despite the fact that the boiler combustion process needs to be optimized to improve the efficiency of coal-fired power generation. Consequently, the temperature and speed of air in the boiler were adjusted to prevent thermal efficiency degradation caused by foreign substances built up on the boiler tube surface and to drastically improve the boiler efficiency [25]. Li et al. utilized and verified a least squares, fast learning network (LSFLN), which demonstrates outstanding performance, even in nonlinear systems that are fairly complicated, inertial, and have a time delay, for a more accurate prediction of the combustion efficiency of a coal-fired boiler [26]. Suntivarakorn et al. constructed an automatic combustion control system using a fuzzy logic control algorithm to execute combustion flue gas heat usage, combustion air preheating, and air–fuel ratio control to obtain boiler efficiency improvement [27].

Wang et al. applied boiler combustion optimization to reduce NO_x emissions in a complicated boiler combustion mechanism based on genetic algorithm (GA) modeling of the Gaussian process (GP). The results are similar to the NO_x emission predictions derived through modeling using support vector machines (SVM) [28]. Shi et al. used an ANN to develop a boiler combustion optimization model for coal-fired power plants and generalized the model using CFD, thus predicting thermal efficiency successfully. The researchers used a GA, which is a multi-objective optimization algorithm, to determine an optimal air distribution method to improve and verify the thermal efficiency of a boiler [29]. Niu et al. built a boiler combustion process model using LS-SVM for boiler combustion system optimization and revealed that the boiler combustion efficiency improved by 0.68% through real-time data mining and online optimization [30]. Vieira et al. proposed a model capable of estimating the steam generation efficiency, power generation, and flue gas of a boiler using an ANN and discovered that the flue gas outlet temperature and air pressure were significant parameters for steam generation efficiency and power generation [31]. Table 2 presents an overview of previous studies by category.

Table 2. Overview of previous literature on power plant efficiency improvement and flue gas O₂ prediction.

Category	Methods/Tools Used for Efficiency Improvement	Year	Authors
Conventional efficiency improvement	Exergy analysis	2012	Murehwa et al. [11]
	Flue gas temperature drop	2012	Karri [12]
	Flue gas pressure drop and induced draft fan overhaul	2012	Mandi et al. [13]
	Air pre-heating using recuperator	2012	Hasanuzzaman et al. [14]
	Parametric analysis and simulation model	2017	Ibrahim et al. [15]
	Nonlinear control strategy based on nonlinear backstepping theory	2019	Errami et al. [16]
	System development by integrating PVS, WEGS, and MHGS	2021	Mosobi and Gao [17]
	Modeling through exergy evaluation	2022	Khaleel et al. [18]
Prediction of flue gas	Flue gas O ₂ monitoring using linear model programming	2019	Zaporozhets [19]
	LASSO algorithm, deep belief network, and least squares support vector machine	2020	Tang et al. [20]
	Long short-term memory modeling	2020	Pan et al. [21]
	Artificial neural network and random forest-based soft sensor	2022	Effendy et al. [22]
	Convolutional neural network modeling	2023	Li et al. [23]
Boiler combustion efficiency improvement	Reducing air with a fuzzy logic controller	2005	Santoso et al. [24]
	Non-dominated sorting genetic algorithm and computational fluid dynamics	2014	Liu et al. [25]
	Least square fast learning network	2014	Li et al. [26]
	Auto-combustion control system using a fuzzy logic control algorithm	2016	Suntivarakorn et al. [27]
	Gaussian process and genetic algorithm	2018	Wang et al. [28]
	Artificial neural network and computational fluid dynamics	2019	Shi et al. [29]

Least squares support vector machine	2020	Niu et al. [30]
Artificial neural network	2021	Vieira et al. [31]

1.3.4. Limitations of Previous Research

Conventional methods for power generation efficiency improvement and other efficiency improvement methods in various domains were examined in a literature review, and it was confirmed that combustion flue gas measurements can be predicted using AI for the stable operation of a boiler. In particular, the unstable performance and malfunction of analyzers causes difficulty in optimizing boiler combustion efficiency, and most studies have focused on various algorithms and software development for the replacement of analysis values. Finally, cases of utilizing AI to optimize the combustion efficiency of power plant boilers, which is the main objective of this study, were reviewed. Some studies have been conducted on combustion air control and thermal efficiency improvement through the development of automatic combustion systems and optimal combustion modeling utilization algorithms. However, most combustion flue gas predictions have been limited to O_2 , and boiler combustion efficiency optimization has been studied separately from stable flue gas prediction. Thus, it was challenging to find a case that comprehensively researched the deduction of boiler efficiency optimization points using the predicted O_2 and CO amounts in the combustion flue gas.

The PLSR applied to the model development in this study enables the prediction of flue gas O_2 and CO values, which are essential for boiler combustion efficiency optimization. It is expected that by selecting key variables from sensor data using PLSR, the prediction accuracy can be improved. Consequently, by applying the O_2 and CO prediction models developed through PLSR into the calculation formula for boiler combustion efficiency optimization, the study derived the operating points for O_2 and CO at which boiler combustion efficiency is maximized. ML is the process of training a computer to learn through data and make improvements through experience instead of explicitly programming the computer for learning and improvements [32]. It is extremely challenging to deduce the optimal points and relevance of certain elements through statistical analyses or the intuition of a skilled expert for a generator operated by recording 361 sensing data points in seconds. Therefore, this study aimed to create a model using ML to predict the amounts of O_2 and CO that complement abnormal measurement values and the frequent breakdown of an analyzer that measures the flue gas of a boiler in which various types of fuels are simultaneously burned. Following that, the O_2 and CO prediction models were replaced with the boiler efficiency calculation equation suggested by the American Society of Mechanical Engineers (ASME), PTC 4.0, to deduce and apply O_2 and CO values that maximize efficiency.

2. Research Process

In this section, the overall research process is described, and the main content of each section is summarized.

Section 4 presents the data selected for the development of a model that predicts the amounts of O_2 and CO in boiler flue gas. Variables related to the flue gas, boiler, and power generation output were selected through feature selection, and unnecessary variables, missing data, and outliers were removed.

In Section 5, the ML algorithms applied to the model development are explained and selected by reflecting on the characteristics of the selected data and research outcomes. In addition, the boiler flue gas prediction model (BFG-PM) is explained, along with the training and fine-tuning procedures.

In Section 6, the target value for judging the relevance of the developed model is determined, and whether the O_2 and CO prediction models completed through training and fine-tuning are appropriate for the performance target is examined. The results and performances of the models are summarized.

In Section 7, the completed BFG-PM for combustion efficiency optimization is substituted into the boiler efficiency calculation equation to determine the amounts of O₂ and CO that maximize the boiler combustion efficiency and to systematize the relationships. Accordingly, the boiler efficiency calculation equation is explained, along with the process of substituting and deducing the correlations between O₂, CO, and BFG-PM. Finally, the system configuration and driving procedure are explained.

In Section 8, field applicability is judged based on simulations and relevant results. The application results for improving the efficiency and determining the combustion state of a boiler are verified to deduce efficiency improvement measures. Finally, in Section 9, the BFG-PM developed in this study and the boiler efficiency calculation equation are used to compute the economic effects of optimal combustion control. Figure 3 illustrates the overall process of this study.

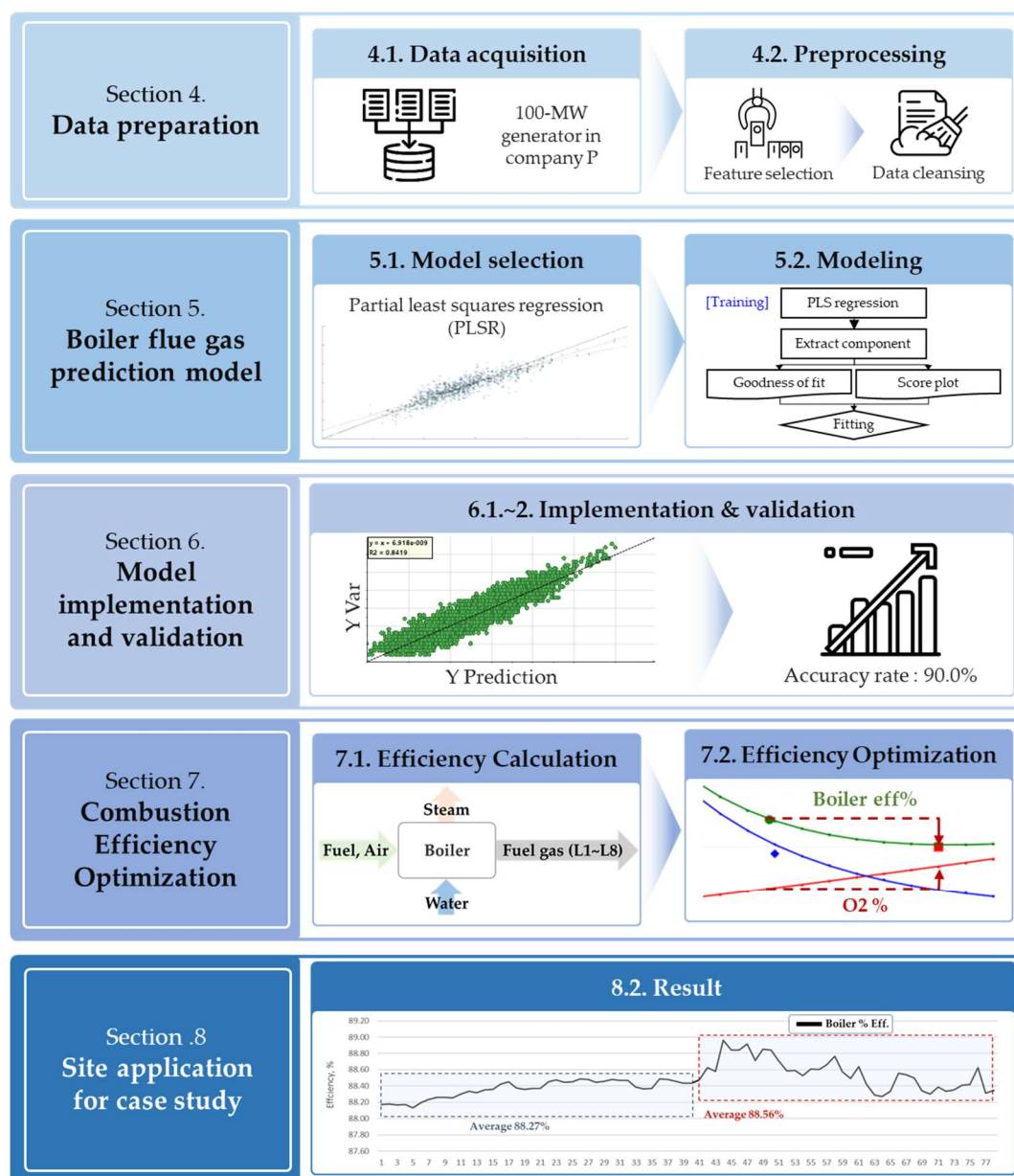


Figure 3. The overall research process.

3. Data Preparation

3.1. Data Acquisition

For this study, operational data for the power generation equipment were collected. The 100-MW generator no. 11, located in Pohang, was selected from among the steam generators operated by company P as the generator from which the data were collected for this study. Generator no. 11 has a high operation rate of electric power generators despite changes in operating conditions at the steel mill, accessible data storage, digitalization through IoT devices, and analyzers for the flue gas O₂ and CO analysis during boiler combustion.

The data collected from sensors in the generator equipment were stored in the operating room PLC and DAQ server and then extracted and stored in the cloud through a manufacturing execution system (MES) of PosFrame, which is a standard platform developed by POSCO DX for long-term storage, analysis, and utilization using P/C and HMI. Figure 4 illustrates the data collection, storage, and extraction processes.

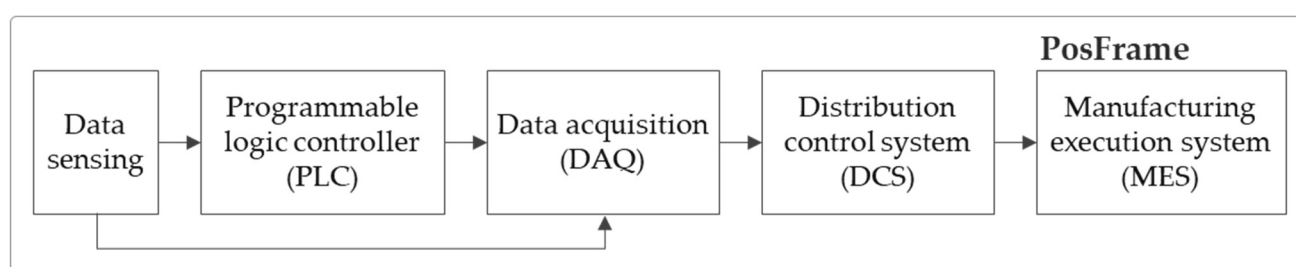


Figure 4. The process of data acquisition from Company P.

The data collection period was one year, from 1 September 2018 to 30 August 2019, to ensure a sufficient amount of time was considered so that the data could be adequately analyzed even if data loss occurred due to the malfunction of analyzers, and the collection interval was set to 10 min. Considering the nature of a generator that is constantly operated, the data type was continuous and structured, and the data did not require analysis in seconds because the conditions did not vary drastically owing to steady operation.

The size of the downloaded data was 116 MB, consisting of 11 classes and 361 features from the raw data of all tags collected during the operation period. The items in the generator data were divided into three categories: boilers, generators, and BOP. Boilers that generate steam by burning fuel are divided into four classes: fuel combustion system boilers, heat-exchange equipment, high-pressure steam boilers, and water supply system boilers. Subsequently, the 167 features were classified and stored as real-time data. A generator that generates power is then divided into two classes of generator and drive systems, and 13 features were classified and stored accordingly. Lastly, BOP is divided into five classes: condenser system, high voltage power equipment, low voltage power equipment, and steam turbine, and then 187 features were stored (Table 3).

Table 3. Data collected from the manufacturing execution system.

Category	Class	Features	Number of Features
Boiler (167)	Fuel combustion system boiler	Including BFG/COG/FOG calorie, BFG/COF/FOG flow control, TDLS O ₂ /CO, flue gas O ₂ , flue gas inlet/outlet temp, total fuel flow, total air flow control	54
	Heat exchange equipment	Including GAH inlet/outlet gas temp, GAH oil pump temp, FSH inlet/outlet steam temp, HPH inlet/outlet fw temp	29
	High-pressure steam boiler	Including FDF wind temp, FDF fan BRG temp, FDF motor BRG temp, AUX process steam temp, AUX steam head temp, STACK GAS flow, STACK_O ₂ , main steam flow, process steam flow	60

Generator (13)	Water supply system boiler	Including platen SH inlet V/V, platen SH out V/V, BFP wind temp, BFP suction FW PH, BFP motor BRG temp	24
	Generator	generator watt, generator mvar, generator zero phase current, gfr fluid oil press	4
	Drive system	Including FDF/IDF VVVF output ampere, FDF/IDF VVVF output voltage, FDF/IDF VVVF RPM	9
	Condenser system	Including BCW CLR inlet/outlet temp, BCW PH, BCW pump outlet pressure, COND overflow control valve, COND pump recirculation control valve	13
	High-voltage power equipment	Including 6.6 kV unit BUS ampere, 6.6 kV unit BUS VAR, ESP unit main TR BCT Ao, main TR wind temp, unit AUX TR current, unit TR oil temp	24
	Low voltage power equipment	Including battery charger DC out current, battery charger battery current, 440 V unit BUS watt	6
	Steam turbine	Including turning speed, turning motor current, turbine inlet steam press, turbine inlet steam temp, cold air exit temp, exciter current, exciter voltage	96
	Water treatment system	including make up water tank level, make up pump ampere, raw water pump out flow, degasifier clear water level	48
	Total		361

3.2. Data Preprocessing

3.2.1. Feature Selection

Feature selection refers to modeling with only significant variables with high explanatory power among given variable candidates, in which the smallest feature subset with a specific generalization error or the top feature subset with the k function is selected to generate a minimum generalization error [33]. Feature selection was performed twice. First, an ideal dependent variable was selected because multiple analyzers were installed in generator no. 11, resulting in six dependent variables related to O_2 . Since the most critical dependent variable for combustion control is O_2 , of the several flue gas O_2 analyzers, data were reviewed based on TDLS_ O_2 _A, TDLS_ O_2 _B, FLUE GAS_ O_2 _A, and FLUE GAS_ O_2 _B. A telemonitoring system (TMS) was installed at the power generation boiler chimney to monitor the emission of air pollutants and had high reliability because it was positioned at the end of the exhaust passage. Accordingly, a correlation analysis was performed between STACK_ O_2 and each analyzer to select the analyzer with the highest correlation. The values of TDLS_ O_2 _B were removed owing to numerous peaks and low reception rates caused by frequent malfunctions. The results of the correlation analysis between the remaining three variables and STACK_ O_2 are shown in Figure 5. Figure 5a shows a 75.1% correlation between the TLDS_ O_2 _A analyzer and STACK_ O_2 ; Figure 5b shows a 2.6% correlation between the FLUE GAS_ O_2 _A analyzer and STACK_ O_2 ; Figure 5c shows a 14.6% correlation between the FLUE GAS_ O_2 _B analyzer and STACK_ O_2 . Therefore, it was determined that TDLS_ O_2 _A in Figure 5a, which had the highest correlation with STACK_ O_2 , demonstrated the most outstanding performance and was thus selected as the main analyzer.

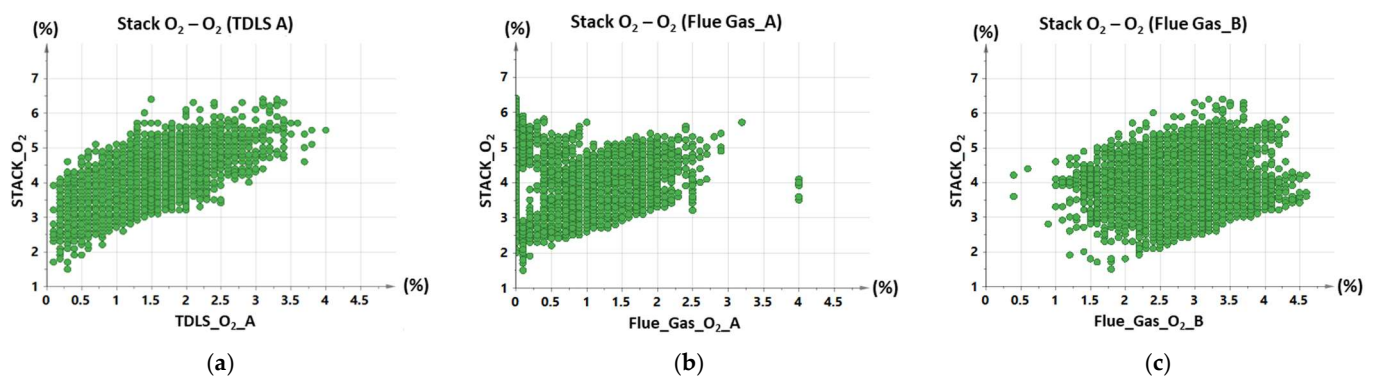


Figure 5. Correlation analysis between O₂ analyzer and STACK_O₂: (a) TDLS_O₂_A analyzer; (b) FLUE GAS_O₂_A analyzer; (c) FLUE Gas_O₂_B analyzer.

Second, executing ML with 356 features involved an excessive number of variables; therefore, preprocessing was performed to select representative variables. For the feature selection of independent variables associated with the combustion efficiency measurement and amounts of O₂ and CO in the boiler combustion flue gas, an analysis was performed using the domain knowledge of four generator operators and two engineers. Consequently, 24 independent variables related to boiler combustion and power generation output were extracted from the fuel combustion system boiler, heat exchange equipment, HP steam boiler, and generator.

3.2.2. Data Cleansing and Derived Variable Creation

Missing data and outliers were removed for data cleansing to prevent distortion and degradation of the learning performance of the model. ‘Missing data’ refers to a state in which no data were recorded, whereas ‘outliers’ refer to values that deviated markedly from the normal range of the observation data [34]. For continuous data, missing data were commonly removed or replaced with mean or median values, whereas outliers were removed or replaced with reference values based on the interquartile range (IQR) or z-score if the observation data did not follow a normal distribution. However, steam power generation at a steel mill involves continuous operation, and all independent variables were considered significant because they were collected based on the changes in the output according to the fuel supply and the changes in conditions according to fuel characteristics. Instead, the data of the reception rate corresponding to 80% or less of the main analyzer with respect to the dependent variables were removed.

Finally, derived variables were generated, in which new variables were created based on the collected data to improve the performance of the analysis. ‘Derived variables’ refers to newly generated variables using existing variables. Derived variables can only be used in a relevant analysis with logical validity because they were subjectively created by an analyst [35]. The fact that the boiler combustion efficiency optimization modeled in this study was highly correlated with the flue gas O₂ amount, which depends on the BFG, FOG, and COG ratios and calories, was discovered through domain knowledge of the operation, and a transformation based on four arithmetic operations and functions was applied to existing variables to generate nine derived variables: BFG%, FOG%, COG%, BFG input capacity%, FOG input capacity%, COG input capacity%, BFG input capacity, FOG input capacity, and COG input capacity. This created a total of 33 independent variables, along with 24 other independent variables.

The initially collected data had 361 features; however, 96.5% were removed through dimension reduction, missing data, outlier removal, and derived variable generation, ultimately leaving 20,697 data points. Figure 6 shows the data pre-processing performed to date.

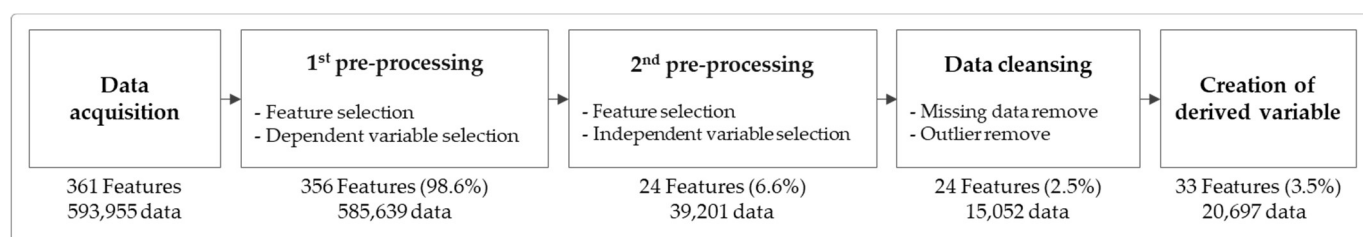


Figure 6. The process of feature selection and data preprocessing.

4. Boiler Flue Gas Prediction Model (BFG-PM)

4.1. Model Selection

An appropriate ML algorithm for the relevant problem must be selected when developing a model. The problem that needed to be solved was whether it belonged to supervised learning, unsupervised learning, or reinforced learning, and whether it was a regression or classification problem. Furthermore, an appropriate algorithm can be selected to determine whether the relationship between the predictive variables and labels is linear or nonlinear, along with an understanding of the characteristics, strengths, and weaknesses of the data [36].

Regression modeling was employed for model development in this study because a dependent variable, Y , exists for an independent variable, X , where data are utilized for modeling to predict the dependent variable, Y , corresponding to a new independent variable, X . In machine learning-based regression modeling, various algorithms exist, including linear regression, logistic regression, K nearest neighbors (KNN), support vector machine (SVM), and PLSR, among others. The choice of algorithm depends on the specific problem and dataset characteristics. It is noted that linear regression has limitations, particularly in terms of predicting accuracy when applied to feature selection, especially when dealing with a large number of variables. Logistic regression, KNN, and SVM algorithms are more suitable for classification. In this study, the PLSR algorithm, which can effectively handle regression tasks when dealing with multi-variable and feature selection problems, was determined to be the most suitable choice. Table 4 presents the characteristics of five popular regression algorithms for supervised learning.

Table 4. Five popular regression algorithms used in supervised learning.

Category	Types	Decision	Characteristic	Advantages	Disadvantage
Linear regression	Regression	Linear	Finding best straight line	Simple model easy /Implementation and interpretation	Poor prediction on non-linear
Logistic regression	Classification	Linear	Binary classification /Categorical prediction /Comparison of other models	Simple model /easy implementation and interpretation	Poor prediction in non-linear
¹ KNN	Regression/Classification	Non Small outlier	Identification a new data	Easy multi- Classification /Intuition and Simple	Vulnerable to outliers /Slow on big data
² SVM	Regression/Classification	Linear/ Non-Linear	Classification of 2 or more groups	Categorical numerical prediction /Low impact on error data, Less overfitting	Multiple combination tests required /Slow to learn /Difficult interpretation

³ PLSR	Regression/Classification	Linear	Considers input and output variables together	Easy control of multi-collinearity /Variable reduction with correlation /Efficient model construction	Difficult interpretation of extracted variables
-------------------	---------------------------	--------	---	---	---

¹ KNN: K-nearest neighbors, ² SVM: Support vector machine, ³ PLSR: Partial least squares regression.

The review of the data characteristics in this study has highlighted several key points. Firstly, there are a large number of variables, and both input and output variables need to be considered simultaneously. Additionally, it has been recognized that there is a need for a regression analysis algorithm capable of variable reduction by examining inter-variable correlations. As a result of these considerations, the study applied the PLSR algorithm to develop the prediction models for boiler flue gas O₂ and CO. PLSR involves extracting k linear combinations with a high covariance with dependent variables. This is frequently used in modeling to find a correlation between independent and dependent variables and consistently uses the least squares method for parts that cannot be explained by the extracted variables [37].

Furthermore, it is also frequently used when there are more predicted variables than observation values and to handle multicollinearity when there are numerous, highly correlated, independent variables. The correlation between X and Y is included in this component because both X and Y are initialized through decomposition [38]. To apply PLSR, multicollinearity was prioritized because a high number of independent variables were highly correlated in regression modeling. The regression model was not significantly affected by the amount of data, was the most basic and highly accurate and estimated the predictive values of dependent variables. Because multicollinearity causes distortion of the statistical significance of an independent variable, thereby inducing an unstable regression coefficient such that analysis cannot be carried out, a correlation analysis was performed to identify the characteristics between features. As a result, at least seven groups had a value of 0.8 or higher between -1 and 1 , which represents the linear relationship of features, thus confirming the high multicollinearity of features. In general, 0.7 was chosen by rule of thumb in the correlation matrix analysis. First, dimension reduction must be performed to solve the multicollinearity problem to ensure that the applied modeling does not cause distortion and obtain a stable regression coefficient. The PLSR algorithm can predict dependent variables through feature extraction, where variables are extracted through the transformation of predicted variables, even if strong multicollinearity is present between independent variables.

4.2. Modeling for Flue Gas Prediction

Modeling is the process of generalizing data patterns by applying specific algorithms after preprocessing and analyzing a prepared dataset [39]. An ML algorithm appropriate for the prepared dataset and the deduced values was selected, and an optimal model was created through training, fine-tuning, and relevance evaluation. The PLSR algorithm, which was selected for predicting the O₂ and CO amounts in boiler flue gas, was applied to the modeling using the SIMCA 13.0.3.0 program developed by SATORIUS. Figure 7 shows the overall process of BFG-PM modeling.

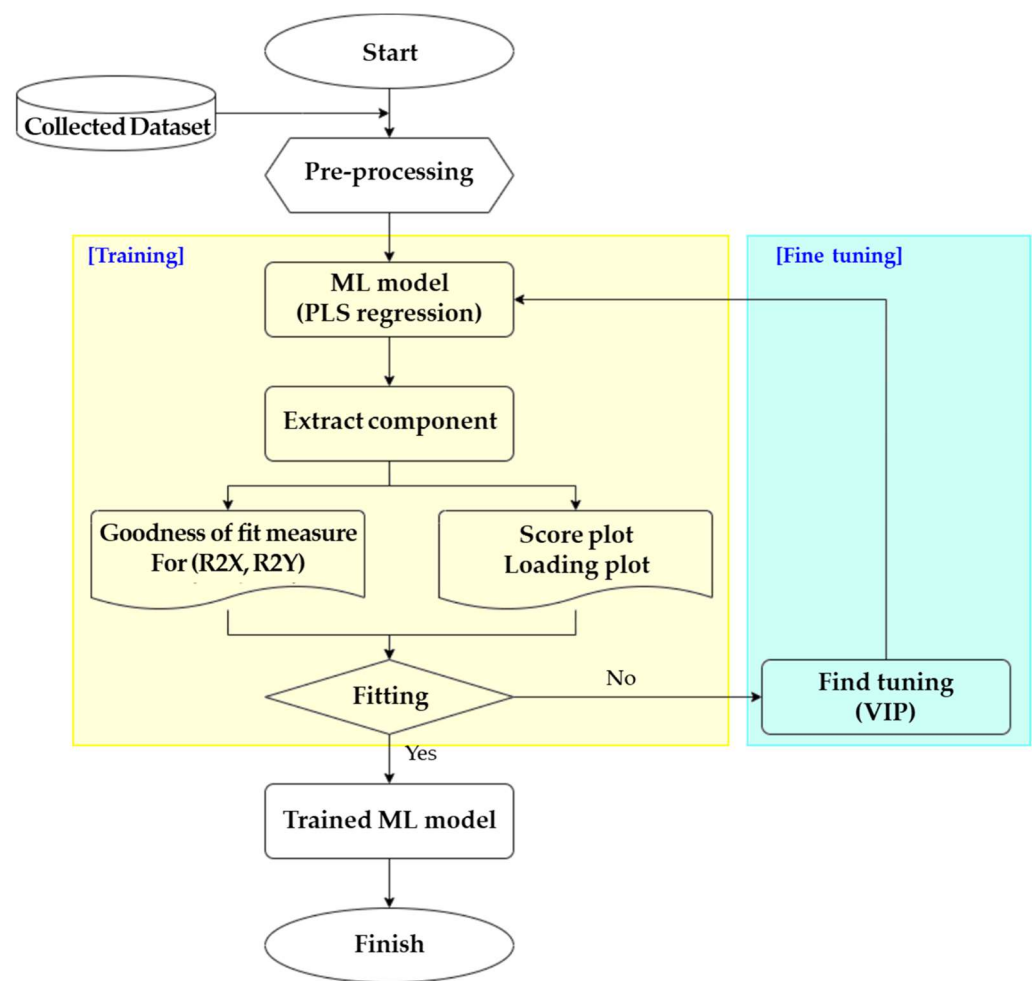


Figure 7. Methodologies and modeling process.

In the PLSR process, which is the core element of the model, one or more variables were generated as components when the initially prepared independent variable data were used as the input. The R^2X for the independent variables and R^2Y for the dependent variables were computed for multiple components generated using the extracted variables, where the R^2X and R^2Y values represent the relevance of the model. A greater R^2X value indicates excellent dependence between variables, whereas an R^2Y value closer to 1 indicates high model reliability [40]. The score and loading plots were examined using a visual search. The score plot represents the extent to which the trained data were within the normal range and outliers were distributed, whereas the loading plot represents the correlation between influential factors with respect to Y . If the model trained in this way is judged to be suitable through the confirmation of the research criterion, model training is completed; otherwise, fine-tuning is performed and repeated using PLSR.

Variable importance in the projection (VIP) was applied for fine-tuning. VIP involves leaving only significant variables and reducing the remaining variables, in which VIP is calculated by multiplying the square of variance and the correlation between the latent variable and the original variable. If VIP is one or greater, it is considered important and not eliminated [41]. Equation (1) represents VIP.

$$VIP_k = \sqrt{K \sum_{n=1}^N SS_n w_{nk}^2 / \sum_{n=1}^N SS_n}, \quad (1)$$

where

K is the total number of signal variables

w_{nk} is the weight of the k th variable for the n th PLSR component

N is the total number of PLSR components

SS_n is the sum of squares explained by the n th PLSR component.

When executing VIP, only variables with high significance were selected from the initially applied independent variables, X , and were then reapplied to PLSR.

5. Model Implementation and Validation

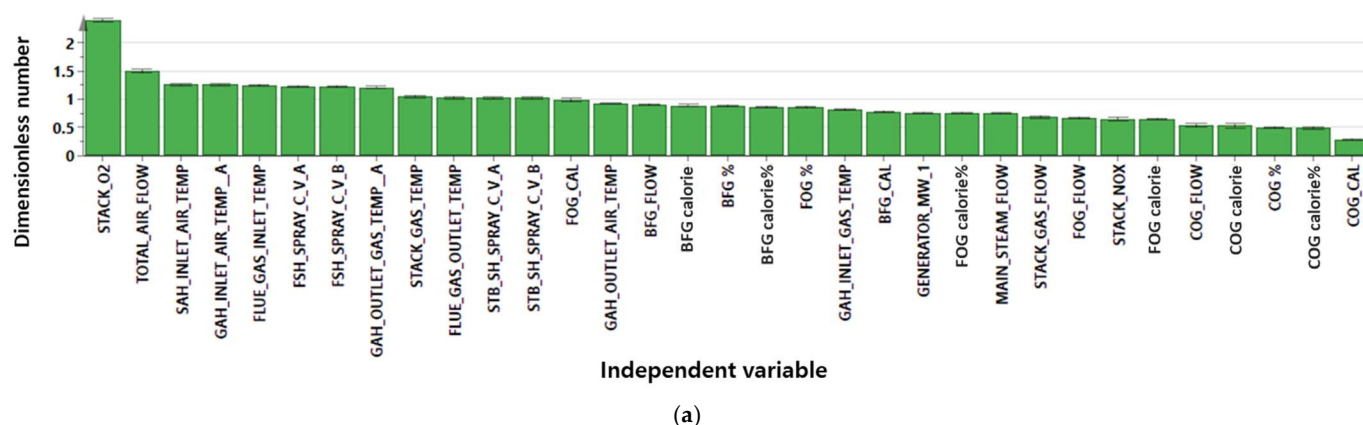
Model training refers to repeating and generalizing the process of finding the optimal parameters to learn various types of data and improve the results [42]. Modeling using the PLSR algorithm creates a model and provides prediction accuracy based on the given variables. Therefore, the final model is determined and verified through training (which maximizes accuracy) and by adjusting the variables based on the model's prediction accuracy.

5.1. Model Implementation for Boiler Flue Gas O_2

5.1.1. Training of BFG-PM for O_2

For training the BFG-PM model for O_2 , three months of data from 18 October 2018 to 11 January 2019, with the least amount of missing data on O_2 , were used. After completing the model, relevance was evaluated based on $\pm 0.25\%$, where the error between measured and predicted values of O_2 was within 10%, reflecting the opinions of field operators and engineers.

For the first training, VIP was computed for all 33 independent variables to identify their priorities and was then applied to the PLSR to examine the variable correlation (R^2X) and goodness-of-fit (R^2Y) according to the components, in addition to the score plot and loading plot. Figure 8a shows that STACK_ O_2 had the largest value, thus having the most significance when determining variable importance based on the VIP of the 33 variables, whereas COG_CAL has the smallest value, thus having the least significance. Fine-tuning must be performed if the model's accuracy using variables is low and the variables to be applied to the subsequent training are selected based on importance. Figure 8b is a loading plot, and it reveals that the dependent variable TDLS_ O_2 _A is positioned in the first quadrant, indicating a positive correlation with other variables. This suggests a high degree of relevance between TDLS_ O_2 _A and the variables in that quadrant. Figure 8c is a score plot, which is used to classify normal data within a 95% confidence interval of the model and identify abnormal data that fall outside this boundary. The results show that there are many abnormal data outside the ellipse, which represents the normal category. This indicates that there are a significant number of data that deviate from the expected pattern and can be classified as abnormal. Figure 8d shows the visual exploration of the relevance between the analyzer and the prediction model, where a scatter located closer to the linear regression equation generated by the model represents high accuracy.



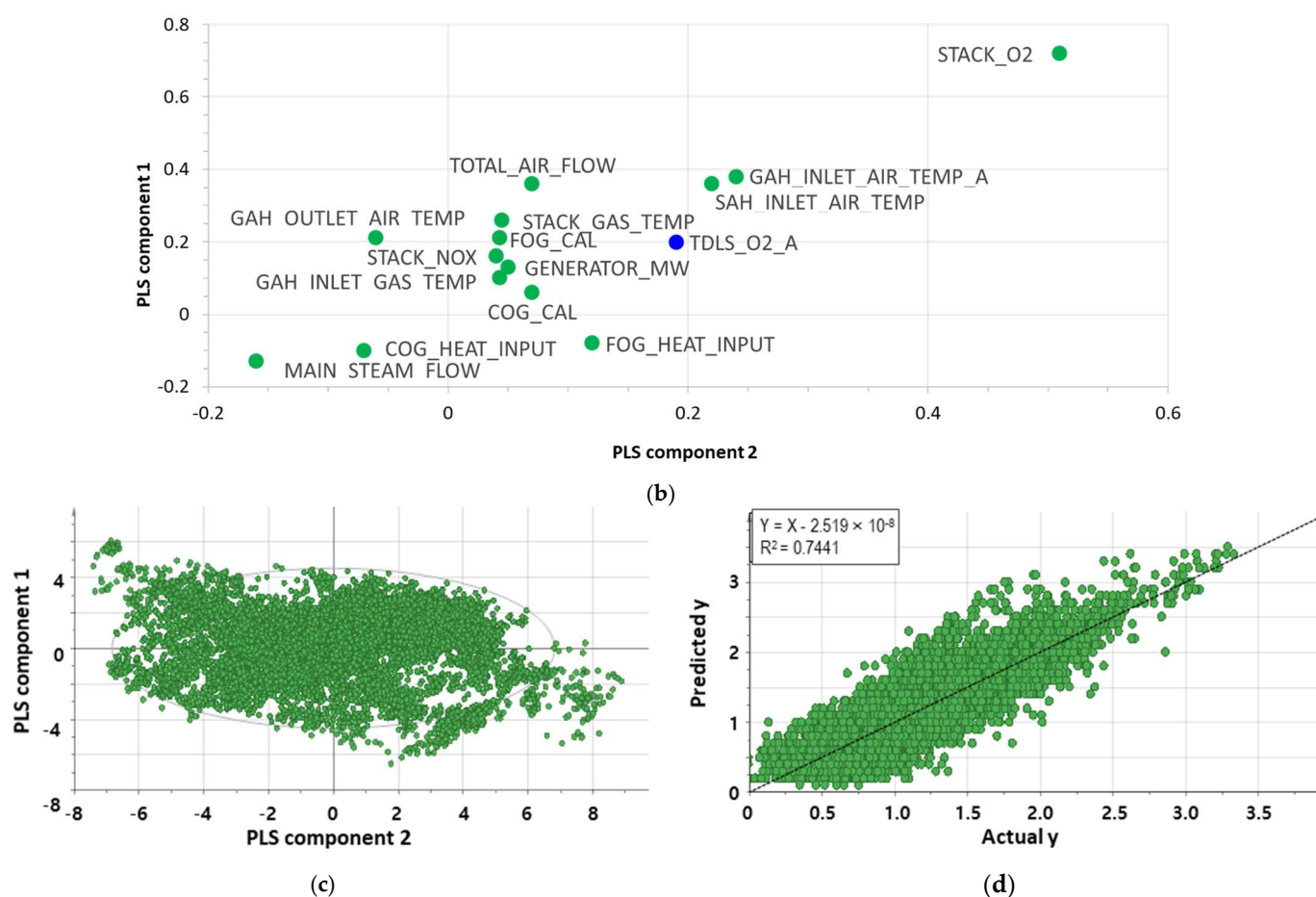


Figure 8. Training output of BFG-PM: (a) VIP of variable 33 ea; (b) Loading plot; (c) Score plot; (d) Match between actual value and prediction value.

When the PLSR was performed using the selected variables, a latent variable, or component, was automatically generated by extracting the variable with a high covariance with the dependent variables. In the first training, the model's reliability was highest when six components were generated, and the process was terminated. The deduced model had explanatory powers of 0.867 for variable dependence (R^2X) and 0.744 for model reliability (R^2Y). Table 5 shows the component creation and goodness-of-fit according to the variable extraction.

Table 5. The variable suitability results based on model training.

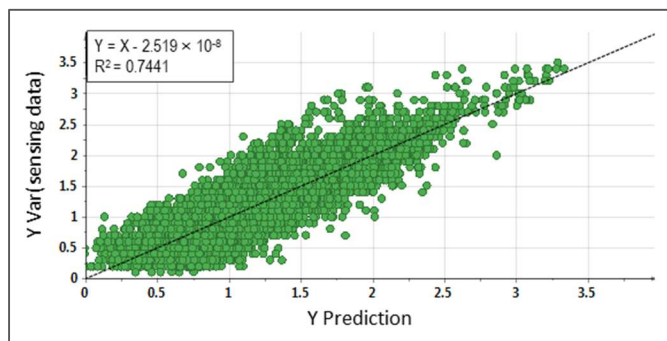
Component	R^2X	R^2Y
1	0.283	0.282
2	0.469	0.413
3	0.524	0.657
4	0.681	0.690
5	0.806	0.721
6	0.867	0.744

Because the model generated using 33 variables had high multicollinearity, a second training session was performed to examine how the model's goodness-of-fit changed when different variables were selected. Variable reduction was performed based on domain knowledge and variable importance using VIP. Spray variables were distinguished into A and B, but variable A was selected as the primary variable because the two values were identical. Input capacity was chosen as the primary variable for fuel variables among

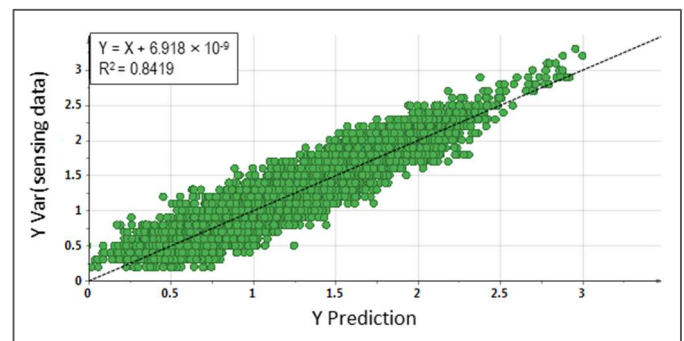
Flow, Cal, %, and input capacity. The power output was excluded because it belonged to the result value. Consequently, a total of 17 process variables related to O₂ were selected. When the model was re-analyzed based on the newly chosen variables, R²X was 0.925, and the model's goodness-of-fit (R²Y) was 0.842.

In the third training session, the model was analyzed using 13 variables, excluding four low variables with VIP, and an R²X of 0.764 and an R²Y of 0.836 were obtained. In the fourth training, the model was analyzed using 12 variables after excluding five low variables, including COG input capacity, in the 17 variables used in the second training; R²X and R²Y were 0.930 and 0.835, respectively, thus demonstrating no significant impact on the model. In the fifth training session, the model was analyzed based on the VIP using 11 variables after excluding certain variables to examine the effects of spray variables: R²X and R²Y of 0.976 and 0.832, respectively, were obtained.

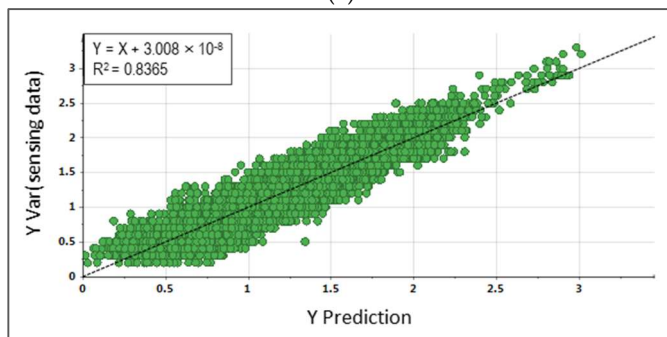
In the sixth training, the model was analyzed based on the VIP using nine variables after excluding the GAH inlet air temperature, which was identical to the SAH inlet air temperature. R²X was 0.974 and R²Y was 0.833. In the seventh training session, the model was analyzed using eight variables after excluding STACK_GAS flow through discussions with and knowledge of field experts. R²X was 0.992 and R²Y was 0.832. The model was analyzed after excluding the FOG input capacity in the eighth training session, as in the seventh training session. R²X and R²Y were 0.945 and 0.803, respectively. Specifically, the R²X and R²Y values degraded abruptly; thus, additional training was terminated, and the seventh training was selected as the final model. Figure 9 shows the reliability obtained in eight training sessions of the BFG-PM model for predicting O₂ in boiler flue gas. Figure 9a shows the initial training results. In contrast, Figure 9h shows the results of the eight training sessions, where the *x*-axis represents O₂ prediction, and the *y*-axis represents the measured value. The functional equation of the predictive model is shown in the top-left corner, where R² indicates prediction relevance.



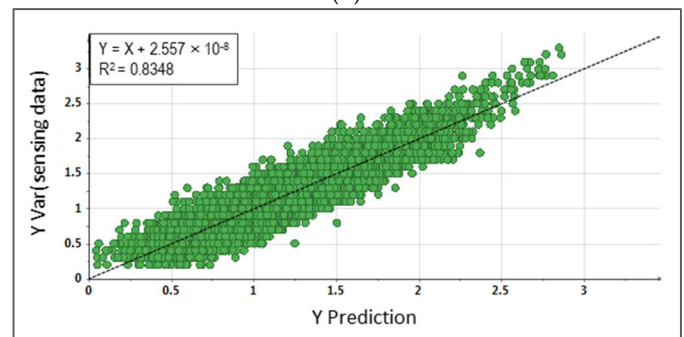
(a)



(b)



(c)



(d)

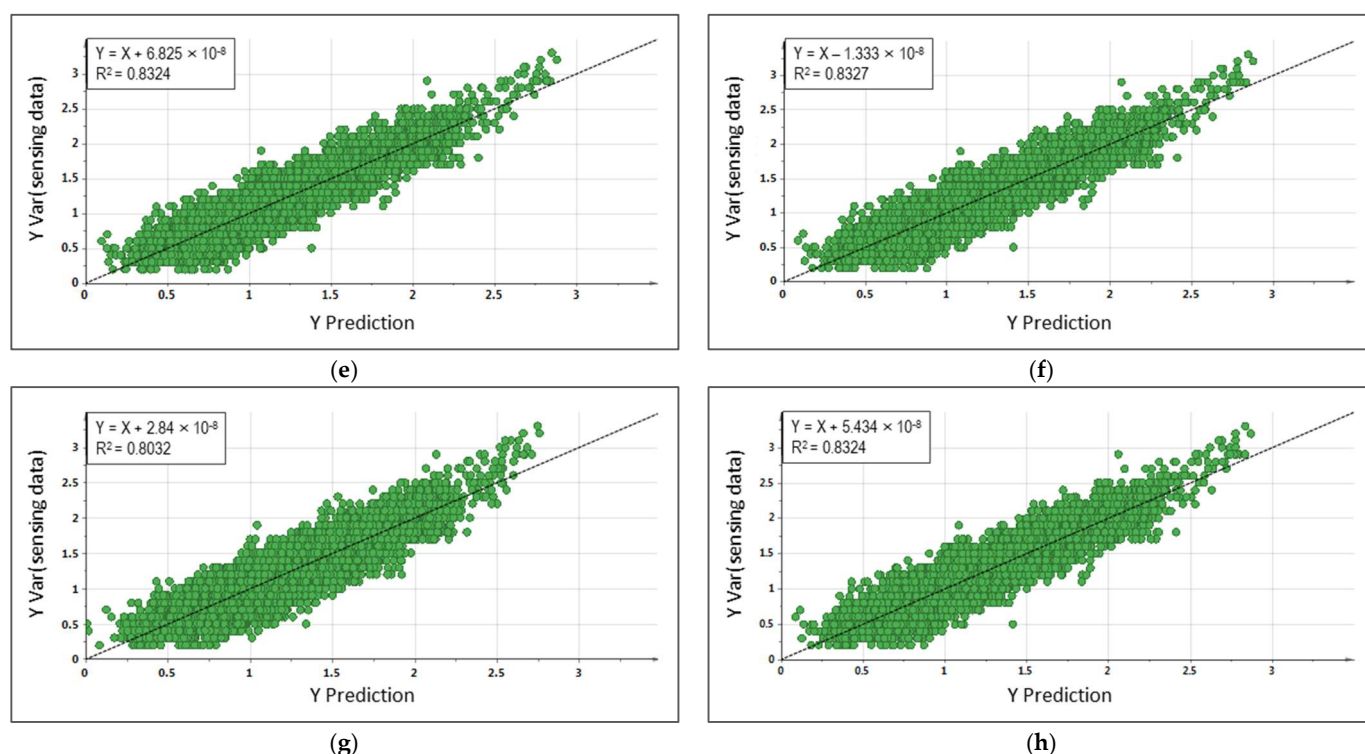


Figure 9. Adaptation of variable Y according to the number of trainings: (a) 1st training - 6th component; (b) 2nd training - 5th component; (c) 3rd training - 4th component; (d) 4th training - 5th component; (e) 5th training - 6th component; (f) 6th training - 6th component; (g) 7th training - 4th component; (h) 8th training - 6th component.

5.1.2. Implementation and Validation

The training results were examined to select the final O₂ prediction model. Variable reduction, including fine-tuning, was performed eight times to improve relevance for training. When the model's goodness-of-fit was analyzed through variable reduction, the reliability remained consistent until the seventh training and then abruptly degraded in the eighth training. Table 6 summarizes the R²X and R²Y according to the number of training sessions. Independence between variables was the most appropriate in the seventh training, whereas the model's reliability was the highest in the second training. To prevent performance degradation due to a correlation between independent variables caused by the large number of variables, the model was selected from the training results involving fewer than 13 variables. Therefore, the final model was set at the seventh training, where R²X was the highest; R²Y remained reasonably consistent after the third training.

Table 6. The training results of the O₂ model.

Category	Variables	Components	R ² X	R ² Y	Model Selection
1st training	33	6	0.867	0.744	
2nd training	17	5	0.925	0.842	
3rd training	13	4	0.764	0.836	
4th training	12	5	0.930	0.835	
5th training	11	6	0.976	0.832	
6th training	9	6	0.974	0.833	
7th training	8	7	0.992	0.832	*√
8th training	7	4	0.945	0.803	

*√: selected model for O₂ prediction.

The final model for predicting O₂ using the BFG-PM included eight variables: SAH_INLET_AIR_TEMPERATURE, GAH_INLET_GAS_TEMPERATURE, GAH_OUTLET_GAS_TEMPERATURE_A, STACK_O₂, MAIN_STEAM_FLOW, BFG HEAT INPUT, FOG HEAT INPUT, and TOTAL AIR FLOW, after seven training sessions, and had an X goodness-of-fit of 0.992 and Y goodness-of-fit of 0.832.

For judging the relevance of the final BFG-PM O₂ model, the data within or outside $\pm 0.25\%$ of error in the O₂ amount were distinguished between the measured values and predicted values of the seventh model, with a reliability of 83.2%. As a result, 90.89% of the predicted values, or 5199 of the total 5199 ea, were within the standard. Table 7 shows the performance of the seventh BFG-PM model for O₂ prediction.

Table 7. O₂ prediction result of BFG-PM.

Category	O ₂ Prediction	
R ² Y	0.832	
Acceptance criteria	O ₂ gap between actual and prediction $\pm 0.25\%$	
Result	90.89%	
Within standard	5199 ea	90.89%
Out of standard	521 ea	9.11%
Total	5720 ea	100%

O₂ prediction using the BFG-PM matched the measured values (83.2%). However, even if the two values do not fit perfectly, the prediction value was 90.89%, which was less than the O₂ gap $\pm 0.25\%$ or the requirement for appropriate operation. Thus, boiler combustion control was determined to be feasible using the predicted O₂ value.

5.2. Model Implementation for Boiler Flue Gas CO

5.2.1. Training of BFG-PM for CO

To train the BFG-PM model for CO, three-month data from 31 May to 31 August 2019 were used. The relevance of the model was judged on CO gap ± 350 ppm, where the difference between measured and predicted values was within 10%, and the training was performed in the same manner as the O₂ prediction model.

For the first training, VIP was computed for all 33 independent variables to identify their priorities and then applied to the PLSR to examine the variable correlation (R²X) and goodness-of-fit (R²Y) according to the components, in addition to the score and loading plots. The variable importance of the 33 variables was judged in the same manner as that for O₂. When 10 components were generated, the goodness-of-fit of the variables, R²X, was 0.750, while the model's goodness-of-fit, R²Y, was 0.453. Abnormal data were predominant in the score plot, while the correlation between the variables was insignificant in the loading plot.

In the second training, the model was analyzed by selecting 17 process variables related to CO (after removing Spray B based on VIP and domain knowledge) and the power output corresponding to the result. Using input capacity as the primary variable among fuel variables of Flow, Cal, %, and input capacity, R²X was 0.976, and the model's goodness-of-fit (R²Y) was 0.543.

The model's results were deduced by repeating the training and fine-tuning to find the most appropriate X and Y. The eighth model was selected after repeating the training nine times.

5.2.2. Implementation and Validation

The training results were examined to select the final CO prediction model. For training, variable reduction, including fine-tuning, was performed nine times to improve relevance (Table 8). The goodness-of-fit of variable X was excellent in the fourth, fifth, and

eighth training sessions. In particular, the goodness-of-fit of variable Y was outstanding in the fifth and eighth training sessions. Therefore, the eighth model was selected to prevent errors from increasing owing to multiple variables by simplifying the model.

Table 8. The training results of the CO model.

Category	Variables	Components	R ² X	R ² Y	Model Selection
1st training	33	5	0.750	0.453	
2nd training	17	10	0.976	0.543	
3rd training	14	9	0.990	0.536	
4th training	11	9	0.995	0.477	
5th training	13	9	0.993	0.534	
6th training	12	8	0.992	0.532	
7th training	10	4	0.753	0.375	
8th training	11	8	0.993	0.534	*√
9th training	9	5	0.873	0.380	

*√: selected model for CO prediction.

The final model for predicting CO using BFG-PM included 11 variables: GAH_OUTLET_AIR_TEMPERATURE, GAH_OUTLET_GAS_TEMPERATURE_A, FLUE_GAS_INLET_TEMPERATURE, FLUE_GAS_OUTLET_TEMPERATURE, STACK_GAS_TEMPERATURE, STACK_O₂, MAIN_STEAM_FLOW, BFG HEAT INPUT, FOG HEAT INPUT, COG HEAT INPUT, and TOTAL AIR FLOW after eight training sessions, and has a variable independence of 0.993 and model reliability of 0.534.

For judging the relevance of the final BFG-PM O₂ model, the data within or outside ± 350 ppm of error in the CO amount were distinguished between the measured and predicted values of the model, with a reliability of 53.4%. As a result, 90.09% of the predicted values, or 9869 of 10,955 ea, were within the standard. Table 9 presents the performance of the eighth BFG-PM model in predicting the CO.

Table 9. CO prediction result of BFG-PM.

Category	CO Prediction		
R ² Y	0.534		
Acceptance criteria	CO gap between actual and prediction ± 350 ppm		
Result	90.09%		
Within standard	9869 ea		90.09%
Out of standard	1086 ea		9.91%
Total	10,955 ea		100%

The CO prediction using the BFG-PM matched the measured values (53.4%). Compared with the O₂ prediction, the scatter distribution was larger, and the prediction accuracy was less effective. The combustion state for boiler combustion control was determined based on both O₂ and CO values. However, most combustion control supply combustion air is only based on the O₂ value, where the CO value is applied as a secondary auxiliary indicator. Although the predicted values were inadequate, the CO values were utilized in the boiler combustion control in this study because 90.09% of the data were within a 10% margin of error.

5.3. Discussion

The amounts of O₂ and CO contained in the flue gas must be accurately measured to determine the combustion state during the boiler combustion process. Thus, this study developed a BFG-PM using an ML-based PLSR algorithm to overcome the measurement

limitations. The amounts of O₂ and CO in the flue gas were predicted, and the prediction accuracy of the final model was improved through training.

The goodness-of-fit of variable X, computed using the BFG-PM O₂ model, was 0.992, demonstrating a high correlation between the selected variables. Accordingly, the goodness of fit of variable Y had an excellent prediction accuracy of 0.832. The model showed an accuracy of 90.89% with a $\pm 0.25\%$ margin of error for O₂. The goodness-of-fit of variable X in the BFG-PM CO model was 0.993, which was appropriate. However, the goodness-of-fit for variable Y was 0.534, which indicates inadequate prediction accuracy, and the scatter distribution graph of the correlation showed a wide dispersion. This can be attributed to insufficient outlier removal during data preprocessing.

The finally selected BFG-PM CO model showed a large scatter distribution. This is analyzed due to the decrease in accuracy caused by a large amount of data loss during outlier removal. It was challenging to minimize data loss in the preprocessing process of this study. Outlier removal is a critical process for improving the performance of a model, but it can lead to data loss. In particular, sensor data for predicting the condition of power generation equipment often contains a lot of noise, making it more likely that data loss will occur during outlier removal. Further research is needed for minimizing data loss during outlier removal in the future. This may involve improving the performance of outlier removal algorithms or developing methods to compensate for data loss during outlier removal. Alternatively, it is necessary to acquire a large amount of data so that the accuracy is not affected, even if data loss occurs. By acquiring a large amount of data, the decrease in accuracy caused by data loss can be offset.

Considering the nature of generators, unit commitment rarely occurs, and the operation is continuous, where the measured values do not exhibit noticeable fluctuations. However, if there is significant variation in the data, it is more reasonable to inspect the equipment's state because the operating conditions may have changed instead of experiencing noise. Therefore, the high CO values observed during data preprocessing were not removed because they were considered to have been caused by incomplete combustion. The analyzer CO values varied substantially from 30–6250 ppm, whereas the BFG-PM model predicted 30–3400 ppm. The sharp increase in the measured CO value was attributed to the peak owing to noise rather than the effect of incomplete combustion. Furthermore, CO values were not consistent within a specific range but varied across all ranges. Thus, the sensitivity of the predicted values was higher than that of O₂, which fluctuated less. If the operation continued with a CO value of 300–1000 ppm, even in the combustion state, complete combustion was considered to have occurred. However, the minimum and maximum values differed by at least three-fold. It was deemed sufficient to be utilized in combustion control if the model's predicted value was within the proper margin of error, even if the measured and predicted values did not match perfectly. Ultimately, the model demonstrated an accuracy of 90.09% within ± 350 ppm margin of error for CO. The accuracy of the O₂ and CO values is vital in general. Still, the model's prediction performance of 90% was fairly accurate, considering the fact that the most challenging issue was the difficulty in checking the boiler combustion state owing to abnormal values, such as peak values, decreased reception rate, and frequent malfunction of analyzers. Therefore, the values predicted by the BFG-PM can be substituted if the boiler flue gas analyzer breaks down or the reliability is decreased.

6. Combustion Efficiency Optimization

6.1. Select the Boiler Combustion Efficiency Calculation

In Section 7, this study deduced the points at which efficiency was maximized based on the prediction model of O₂ and CO in the flue gas, which are major factors in the boiler combustion control. Thus, the boiler efficiency must be calculated using an in–out method or a heat loss method, both commonly used [43]. The in–output method is relatively simple. Still, it is easily affected by errors in the analyzers and is less accurate. In contrast, the

heat loss method is more complicated but less affected by errors in the analyzers and is highly accurate. Hence, the heat loss method, which is standardized in ASME PTC4.0, was adopted for the efficiency calculation equation for boiler combustion efficiency optimization [44]. Most industries inspect boiler performance using procedures recommended by the ASME [45]. Considering how efficiency loss factors are categorized into eight types, that each loss rate is taken into account when calculating the overall efficiency, and that the method is not directly affected by measurement errors of calorific value and flow rate, it was deemed appropriate for calculating the efficiency of a generator boiler operating in the adverse environment of company P. Figure 10 illustrates the measured elements required for the heat loss method.

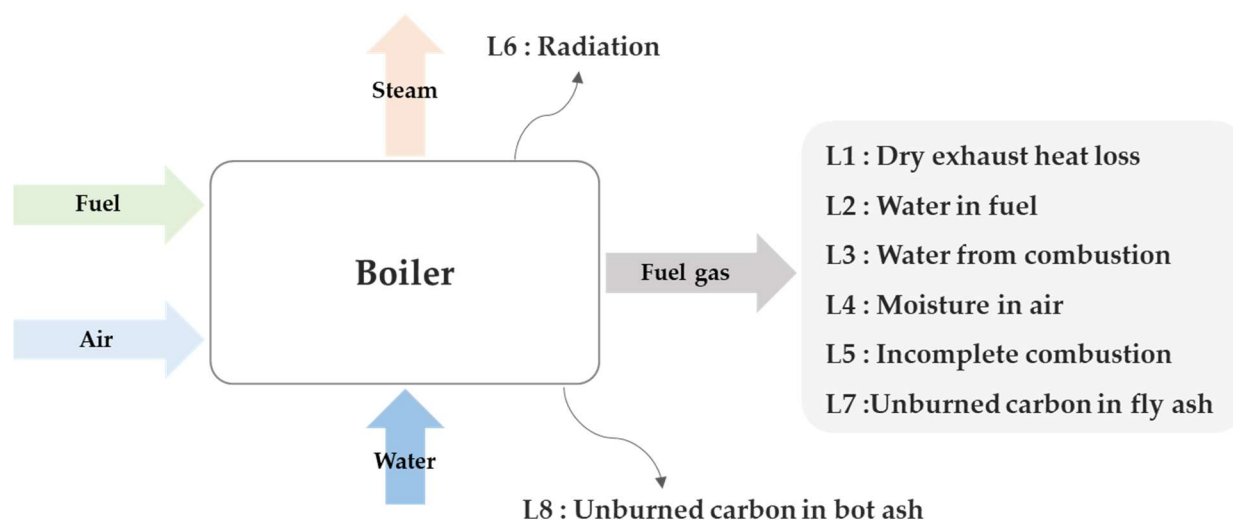


Figure 10. Heat loss method for boiler performance test.

Equation (2) was used for calculating efficiency by reflecting the boiler efficiency loss factor.

$$\% \text{ Eff.} = 100 - (L1 + L2 + L3 + L4 + L5 + L6 + L7 + L8) \quad (2)$$

To calculate the boiler combustion efficiency, 12 variables, including GAH out-gas temperature, O_2 , CO, CO_2 , ambient temperature, humidity in the air, BFG calorie, COG calorie, FOG Calorie, BFG flow, COG flow, and FOG flow, were needed to calculate the heat loss of each item. All variables could be verified and applied through real-time data, but O_2 and CO values were predicted using BFG-PM. Table 10 presents the detailed equations for heat loss calculations.

Table 10. Equations for heat loss calculations.

Loss Calculation	Equations
Loss 1, Dry exhaust heat loss	$((0.981790389084853 \times B12 - 0.82499216637857 \times B13 + 0.905867445657761 \times B14) + (1 + B4/(21 - B4)) \times (0.595123830015147 \times B12 + 11.3028230711785 \times B13 + 1.04479223434223 \times B14)) \times (1.38906921002696 \times B12 + 0.492352550459239 \times B13 + 1.38823214285714 \times B14) \times (B3 - B7) / ((B12 + B13 + B14) \times (B9 \times B12 + B10 \times B13 + B11 \times B14)) \times 23$
Loss 2, Water in fuel	-
Loss 3, Water from combustion	$4.05 \times (0.201568899331099 \times B12 + 20.263286732307 \times B13 + 1.04449389639958 \times B14) / (B12 + B13 + B14) \times (B3 - B7) / ((B9 \times B12 + B10 \times B13 + B11 \times B14) / (1.38906921002696 \times B12 + 0.492352550459239 \times B13 + 1.38823214285714 \times B14))$

Loss 4, Moisture in the air	$(1 + B4/(21 - B4)) \times (59.512383006586 \times B12 + 1130.28230712112 \times B13 + 104.479232322713 \times B14) / (B12 + B13 + B14) \times B8 \times 0.45 \times (B3 - B7) / ((B9 \times B12 + B10 \times B13 + B11 \times B14) / (1.38906921002696 \times B12 + 0.492352550459239 \times B13 + 1.38823214285714 \times B14))$
Loss 5, Incomplete combustion	$B5 / (B5 + B6) \times (18.6804298424681 \times B12 + 43.3664332479575 \times B13 + 27.9736561145342 \times B14) / (B12 + B13 + B14) \times 5744 / ((B9 \times B12 + B10 \times B13 + B11 \times B14) / (1.38906921002696 \times B12 + 0.492352550459239 \times B13 + 1.38823214285714 \times B14))$
Loss 6, Radiation	0.50
Loss 7, Unburned carbon in fly ash -	
Loss 8, Unburned carbon in bot ash -	

6.2. Efficiency Optimization

The fact that boiler efficiency was highly correlated with O₂ and CO values, which indicates the combustion state, is already known, where only O₂ and CO were the predicted values instead of the measured values among the variables used for calculating the efficiency. Therefore, the O₂ and CO values resulting in the highest boiler efficiency could be deduced by applying the O₂ and CO prediction BFG-PM model created using the PLSR algorithm to the given boiler efficiency calculation equations. For this purpose, the two predicted variables for calculating the boiler efficiency were replaced with a single variable to obtain a single variable value where the boiler efficiency was maximized. This process involved replacing CO in the relation equation between the BFG-PM O₂ model and the BFG-PM CO model, and the value was substituted into Equation (2) to deduce the O₂ value at which the efficiency is maximized. The relationship between the two models was replaced with CO because O₂ becomes the first criterion for determining the combustion state. The relationships between the models were identified using the Generalized Reduced Gradient (GRG) method in MS Excel (Equation (3)).

$$Y = 8000 \times 10^{-7.927X} \quad (3)$$

where

Y represents CO ppm

X represents O₂%.

The optimal combustion control value was deduced by linking the relationship between the predicted O₂ and CO values and boiler efficiency. The minimum output of the generator was set to at least 70% operation. The reason for determining the O₂ and CO control values that resulted in optimal combustion was to ensure that the boiler efficiency was optimal by operating the generator with each control value. Consequently, the boiler efficiency was optimal (89.81%) when O₂ was 1.4–1.6%, and CO was 454–372 ppm. Table 11 presents the operating conditions that induce optimal boiler efficiency or predicted O₂ and CO values with respect to the by-product gas usage ratio of BFG 24%, FOG 68%, and COG 7%.

Table 11. Boiler efficiency according to O₂ and CO condition.

O ₂ prediction	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.5	1.6	1.8	2.0	2.2	2.4
CO prediction	1508	1234	1011	827	677	555	454	411	372	304	249	204	167
%Eff	89.43	89.56	89.65	89.72	89.76	89.79	89.81	89.81	89.81	89.80	89.78	89.76	89.73

The boiler combustion efficiency was optimized by considering each variable gathered in real-time in the boiler efficiency calculation equation based on the heat loss method and applying Equation (3) obtained using BFG-PM O₂ and BFG-PM CO. Based on this, a boiler efficiency calculation formula was developed to facilitate easy real-time monitoring of the optimal O₂ and CO values during operation. Building on this foundation, a power-

generation efficiency management system was developed. This system was based on JAVA programming and was embedded in Company P's MES to serve as guidance for optimizing the boiler combustion efficiency.

Figure 11 shows the user interface (UI) of the power generation efficiency management system developed in this study. The current operational state was displayed when generator no. 11 was selected in the UI. The current output, the amount of low-pressure steam produced, and real-time measurements of O₂ and CO were displayed. The O₂ and CO values predicted using the BFG-PM model are provided, along with how the boiler efficiency changes when the current predicted O₂ and CO values are adjusted to the values suggested in the guidelines. The boiler efficiency curvature is shown to help understand this relationship more intuitively. Therefore, it helps determine how to adjust the O₂ value and the extent to which the boiler efficiency is improved accordingly.

When the O₂ value was 1.09% during operation, the boiler heat loss was 11.4%, whereas adjusting the O₂ to 1.40% reduced the boiler heat loss to 10.96%. As a result, the guidance indicates a 1.81% increase in boiler efficiency (Figure 11).

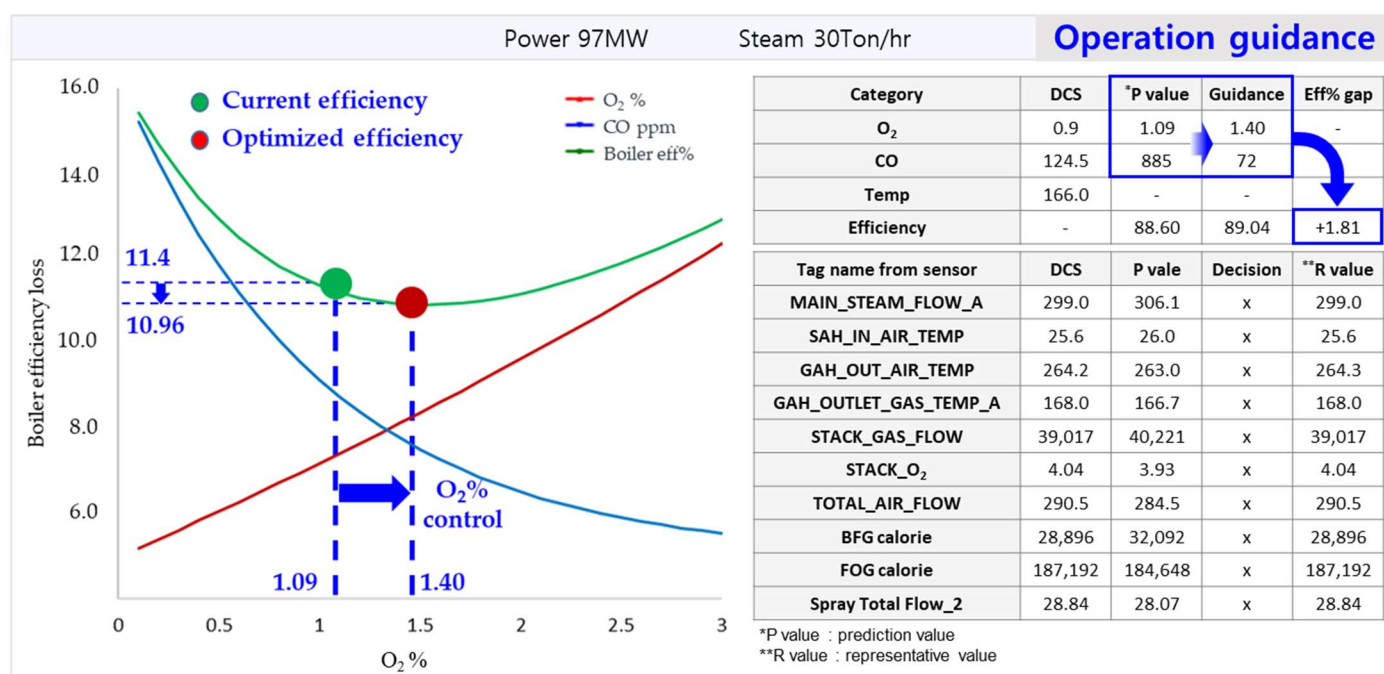


Figure 11. User interface of power generation efficiency management system.

7. Site Application for Case Study

In Section 8, the optimal operation points deduced using the BFG-PM model and the boiler efficiency calculation equation were applied to the equipment in operation to examine whether they operated adequately. Because the BFG-PM was explicitly developed for the characteristics of generator no. 11, the optimal boiler combustion efficiency operation system was applied to generator no. 11.

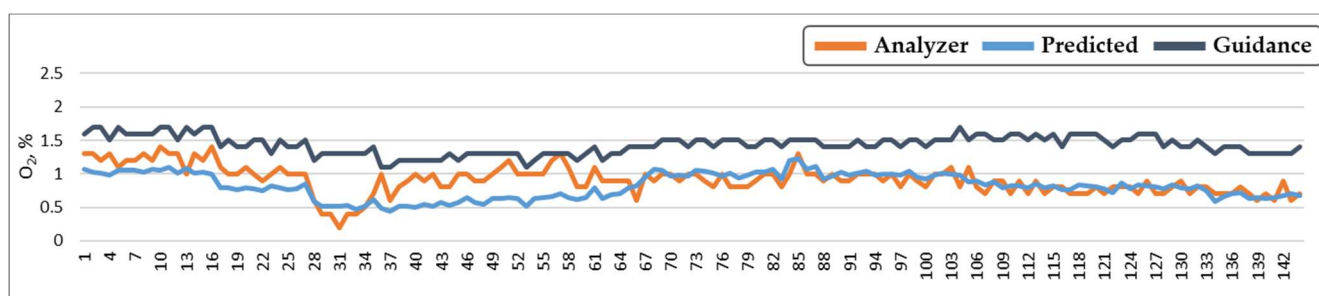
7.1. Simulation for Combustion Control

The simulation was performed using the data of three random days, where the output of generator no. 11 was the highest, and the operation was stable, but it was not used for training and validation of the model. The selected dates were 15 February 2019 and 10 May 2019, when the TDLS O₂ analyzer malfunctioned, and 29 August 2019, when the analyzer operated normally; the data for 24 h were collected every 10 min.

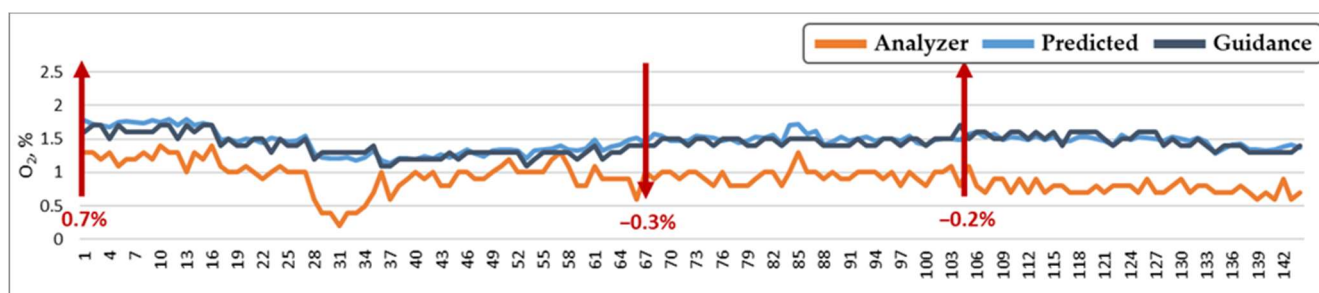
The boiler efficiency can be optimized. First, the similarity between the measured and predicted values was examined using raw data to explore the extent of model completion

visually. If the O₂ operation range that maximized the boiler combustion efficiency was guided to be 1.4–1.6% according to the operation state and data analysis, the predicted values were adjusted accordingly. Finally, whether the predicted O₂ and CO values matched the guidance values was verified according to the O₂ operation range adjustment, during which the boiler combustion efficiency improvement was examined.

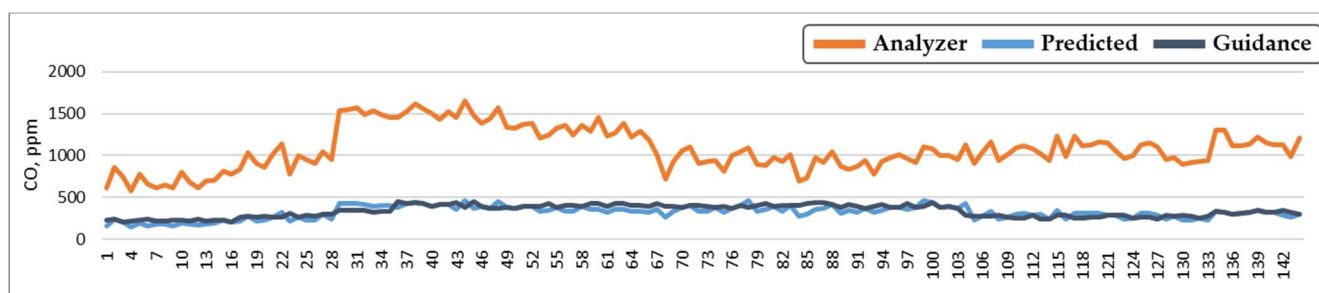
In the first simulation, the goodness of fit of the measured and predicted O₂ values was explored visually. Figure 12a shows the measured, predicted, and guidance values of O₂, where a similarity between the measured and predicted values was observed. In addition, the predicted O₂ value was 0.7% lower than the guidance value for ensuring optimal boiler combustion efficiency. Therefore, the predicted O₂ value increased by 0.7% and then decreased at 670 min, when the predicted O₂ value was higher than the guidance. Finally, it increased by 0.2% at 1040 min when the predicted O₂ value was lower than the guidance value. Figure 12b shows the result of adjusting the predicted value according to O₂ guidance. Figure 12c shows the results of operating the system according to O₂ guidance, where the predicted CO value also followed the guidance, in which the average CO value decreased from 1078 ppm to 318 ppm. Based on the operational guidance, the boiler combustion efficiency was improved by an average of 0.20% and a maximum of 0.37% (Figure 12d).



(a)



(b)



(c)

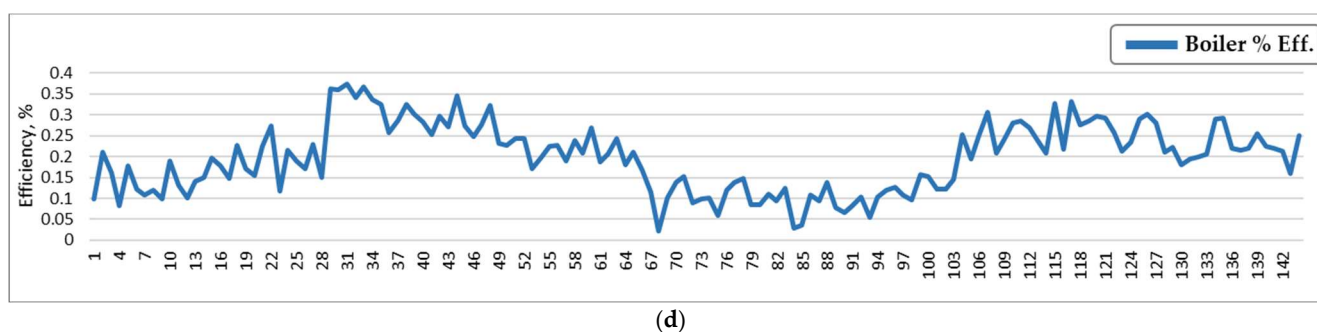


Figure 12. Final simulation test with the data on 29 August 2019: (a) Raw data for simulation test; (b) Adjustment of predicted O₂ value for O₂ guidance value; (c) Changed predicted CO value for CO guidance value; (d) Boiler efficiency increase as O₂, CO optimization.

In the second simulation, the similarity between the measured and predicted values could not be verified because the O₂ sensor malfunctioned. Because an O₂ gap of 0.7% existed between the predicted and guidance values, the predicted O₂ value increased by 0.7%, and the predicted O₂ value was higher than the guidance value at 430 min, decreasing by 0.4%. The value was adjusted to 0.2% at 910 min because the predicted O₂ value was lower than the guidance value. Based on the virtual sensor, the average CO value decreased from 791 ppm to 307 ppm, while the boiler efficiency increased by 0.13% on average and 0.45% at maximum.

In the final simulation, the similarity between the measured and predicted values could not be verified because of a malfunction in the O₂ sensor. Because an O₂ gap of 0.7% existed between the predicted and guidance values, the predicted O₂ value was increased by 0.7%; subsequently, the value was adjusted by 0.2% at 610 min because the predicted O₂ value was lower than the guidance value. Consequently, the average CO value decreased from 772 ppm to 256 ppm, while the boiler efficiency increased by 0.14% on average and by 0.31% at maximum.

A combustion control simulation was performed according to O₂ guidance, maximizing boiler efficiency. Consequently, the CO value matched the guidance value (the optimal operation point) when the predicted O₂ value reached the guidance value. The boiler efficiency increased by 0.13% on average and by a maximum of 0.20%.

7.2. Site Application and Result

The response and accuracy of the boiler combustion control based on the predicted O₂ value were verified using simulation. The site application was carried out under the following conditions: the supply and demand balance was stable because the steel mill consumed less power than the power supply; the power generation output of generator no. 11 was maintained above 70%, while the fuel usage ratio was set to BFG 24%, FOG 68%, and COG 7%. On 13 December 2019, the site application date was selected as it met the required conditions; a zirconia O₂ analyzer was used instead of the TDLS O₂ analyzer, which did not function properly. The application time was one hour, and manual operation was applied to respond to trips caused by an unexpected clash between the system and the generator; a period of one hour was approved by the generator manager, considering a significant loss and the impact on the steel mill operation in the case of abnormality.

Once the site application began, the adjustment of the O₂ value for boiler combustion efficiency optimization was checked. To meet the guidance value of 1.6%, the O₂ value predicted by the BFG-PM was adjusted by +0.3% from 1.3%; the SV adjustment using the zirconia analyzer was enabled at 13 min for combustion control, and the operation mode was changed from remote to manual. Because the O₂ present value of the current analyzer was 2.6%, the set value automatically changed from 2.0% to 2.6% and then remained sta-

ble. The O₂ value needed to be adjusted by +0.3% 20 min after the test started for the optimization of the boiler combustion efficiency; thus, the analyzer O₂ set value was increased by 0.3%, but the present value did not respond immediately and operated nearly at the set value after 10 min. That is, the fuel and combustion air inside the boiler were burned according to the desired O₂ target value. Furthermore, the O₂ value predicted by the BFG-PM was operated at 1.3–1.4%, but optimal combustion control was attained after 10 min when the value reached 1.5–1.6%, similar to the guidance value.

Figure 13 shows the results of the site application tests. When the efficiency was analyzed by applying the system to optimize the boiler combustion efficiency, it improved by +0.29%, from 88.27% to 88.56%, on average. Figure 13 shows the efficiency improvements based on the site-application test.

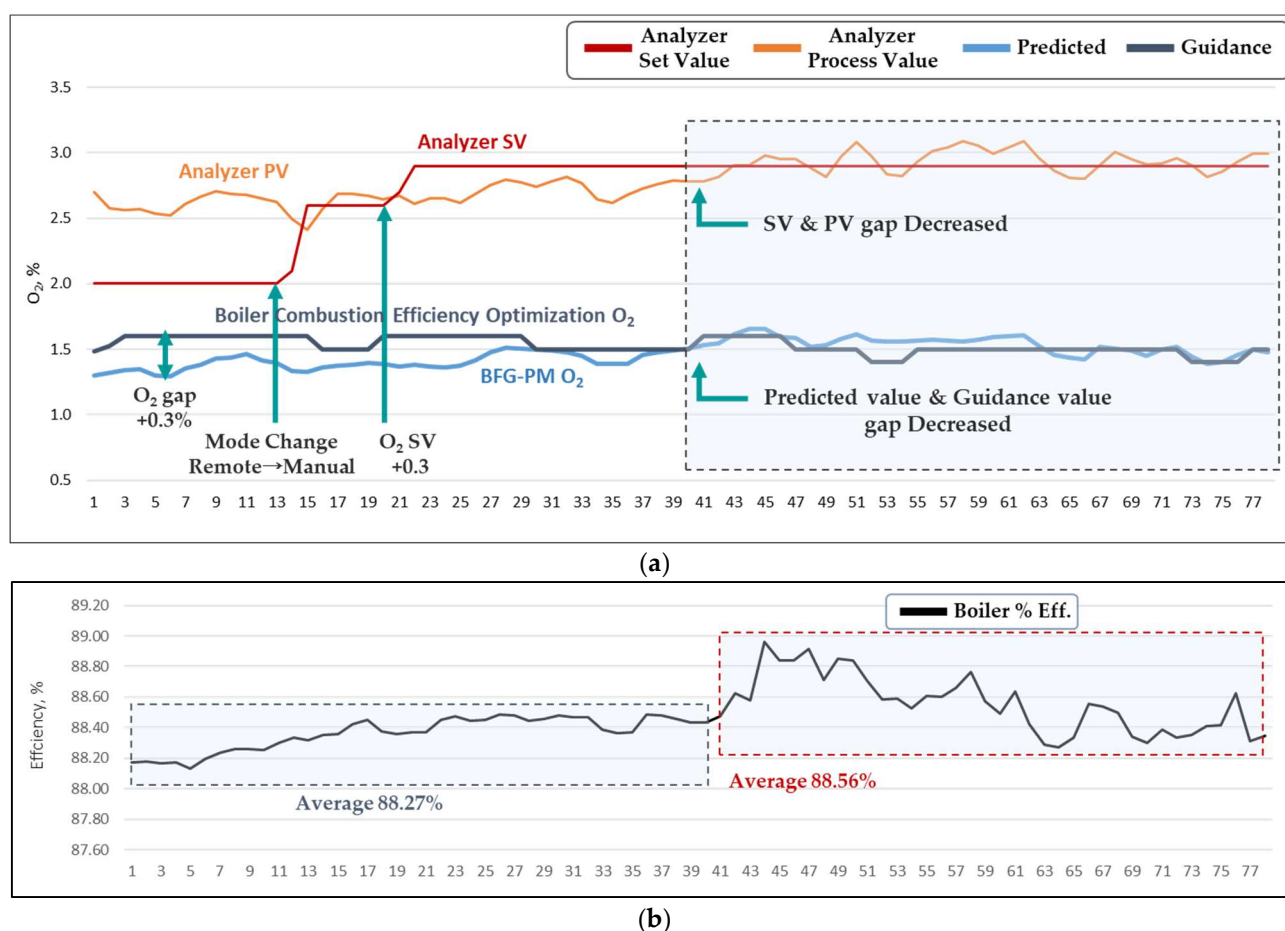


Figure 13. Site application for boiler combustion efficiency optimization with BFG_PM O₂, CO: (a) The process of boiler combustion control with O₂; (b) Change in the boiler efficiency.

This study presents the following distinctive features. O₂ and CO control points were deduced to optimize boiler combustion efficiency and verify efficiency improvements. Theoretically, the maximum and minimum boiler efficiencies differed by 0.38%, depending on the operating ranges of O₂ and CO. However, the efficiency level differed by 0.83% after site application. This is attributable to changes in other conditions, such as an adjustment in the steam production amount or power generation output. This study differs from other studies in that it applies the research outcomes to generators operated at a steel mill rather than quantitatively comparing how much efficiency has improved. Previous studies have been conducted on improving efficiency through boiler combustion control using ML, but no studies have applied the system to power generators.

8. Benefits Analysis

In this section, we calculated the financial impacts of site improvements. This study developed a model for predicting O₂ and CO in boiler flue gas. The efficiency calculation equation was utilized in the boiler to deduce the amounts of O₂ and CO that would enable boiler operation with optimal efficiency. A quantified boiler efficiency improvement was the first observed effect. The overall power generation efficiency improved when the boiler efficiency was improved, whereas the amount of power generated could be increased by using the same amount of adequate fuel. Increased power generation would reduce the amount of power purchased from external sources, thus lowering Company P's external energy purchase cost. Reducing power consumption also leads to a decrease in greenhouse gas emissions. Therefore, quantitative effects were calculated based on the 78-min-long data from 13:19 to 14:36 on 13 December 2019, when the site application took place, to compute the financial impact.

The most noticeable effect is a decrease in power purchase costs resulting from increased power generation output and a reduction in externally purchased power owing to boiler efficiency improvement. The boiler efficiency improved by 0.29% when the O₂ and CO values were set to the guidance values during operation to optimize the boiler combustion efficiency. If the 0.29% efficiency improvement is multiplied by the average power generation amount of 89 MW of generator no. 11 and then multiplied by 24 h, the daily power generation amount increases by 6.2 MWh. The amount of power generated increases by 2161 MWh annually if an average operation rate of 96% over 365 days is reflected. To convert the reduction in the amount of received power to a reduction in purchase cost, the average monthly expense and the hourly charge of industrial power at high pressure among the standard charges of the Korean Electric Power Corporation (KEPCO) in January 2023 were reflected as the unit cost of received power [46]. The cost of receiving 1 kWh of power was USD 0.1 on average. If this amount is reflected in the amount of increased power (+2196 MWh), approximately USD 217,000 (KRW 2.74B) could be saved annually (as of 1 January 2023).

The financial impact of reducing greenhouse gas emissions was calculated by multiplying the annual amount of increased power generation due to efficiency improvement by the greenhouse gas emission coefficient and then multiplying the result by the unit price of emission trading. CO₂eq, which includes CO₂, CH₄, and N₂O, was used as the greenhouse gas emission coefficient of the generated power, whereas the emission market closing average in 2022 was applied as the unit price of emission trading. Consequently, approximately USD 19,700 (KRW 0.25B) could be saved annually. Table 12 presents the results of calculating the quantitative effects based on the operational status data of generator no. 11.

Table 12. Cost calculation for benefit (as of 1 January 2023).

Category	Calculation Result
Energy cost	$89 \text{ MW} \times 0.29\% \times 24 \text{ h/day} \times 365 \text{ day/year} \times 96\% \times 1000 \text{ kWh/MWh} \times \text{USD } 0.1/\text{kWh} = \text{USD } 217,000/\text{year} \text{ (KRW } 2.74\text{B/year)}$
The cost of emission trading	$89 \text{ MW} \times 0.29\% \times 24 \text{ h/day} \times 365 \text{ day/year} \times 96\% \times 0.4781 \text{ tCO}_2\text{eq/MWh} \times \text{USD } 19/^* \text{ tCO}_2\text{eq} = \text{USD } 19,700/\text{year} \text{ (KRW } 0.25\text{B/year)}$

* ton of CO₂ equivalent.

9. Summary

9.1. Conclusions and Contributions

This study used ML data analysis and deduced systematic improvements to reduce the cost of purchasing power from external sources by increasing the power generation output and optimizing the boiler combustion efficiency at company P. The improvement target for increasing the power generation output was limited to the boiler, and boiler efficiency was improved through combustion control. First, a total of 593,955 data points, in which 361 features were stored for a year, were extracted to develop the model, the

extracted data were preprocessed, and derived variables were generated to ensure that the data were more appropriate for modeling. Subsequently, the PLSR algorithm was selected, which was suitable for data characteristics and target values. Training and fine-tuning were carried out to choose the final variables. As a result, the BFG-PM model appropriate for O₂ and CO predictions was created. The final model accurately predicted more than 90% of the data within a 10% error margin, as the relevant standards suggested. Therefore, the predicted O₂ and CO values were considered reliable. The boiler combustion efficiency was optimized using the BFG-PM model, which was developed to replace the inaccurate sensing of boiler flue gas O₂ and CO analyzers.

A boiler efficiency calculation equation was required, for which the commonly used ASME PTC4.0 was applied. The O₂ and CO values at which boiler efficiency was maximized were deduced by applying the O₂ and CO model relationship equation to the efficiency calculation equation. Data were provided on how much the boiler efficiency improved if the O₂ value was adjusted according to the guidance when the boiler was operated based on the current boiler flue gas O₂ and CO predicted values. To examine the changes in efficiency when the boiler was operated according to the guidance, a power plant boiler in an online state was tested at the site for one hour; as a result, the boiler efficiency improved by 0.29%, and the power generation output increased by 258 kW.

This study makes the following contributions. First, ML was used to predict the O₂ and CO concentrations in the flue gas and optimize the efficiency of a power plant boiler. Second, the flue gas O₂ and CO values at which the power plant boiler operated with high efficiency were examined with respect to changes in the flow rate and calories of the power-generation fuel. Finally, the models developed in this study and the optimized boiler efficiency points were systematized and applied in the operational field, utilizing them as guidance to enhance the efficiency of boilers during operation.

Additionally, this study contributes to standardizing boiler operation theory. By applying the developed models to the operational field and calculating their economic impact, this study provides valuable information that can contribute to energy conservation and cost reduction. Moreover, this study has implications regarding the financial benefits resulting from the increased efficiency of power generator 11 and qualitative senses. The reduction in greenhouse gas emissions through improved power generation efficiency is a non-financial element of a company; rather, it corresponds to environmental, social, and governance (ESG) management practices for the sustainable development of a company. Accordingly, the company's brand image can be enhanced. By automating the boiler flue gas O₂ control, the workload of the operator is reduced while improving the operational efficiency of a junior operator. The produced effects will be more excellent if the BFG-PM model is designed according to the characteristics of other steam generators and if boiler combustion optimization is expanded even further. Multiple sensors need not be installed to measure O₂ and CO in the boiler flue gas accurately, and the number of O₂ and CO analyzers installed at multiple combustion facilities operating at company P can also be minimized.

9.2. Limitations and Future Works

In this study, a boiler flue gas O₂ and CO prediction model was developed, and the boiler combustion efficiency optimization points for generator no. 11 were deduced and verified. However, there are certain limitations in creating a flue gas CO prediction model and its accuracy. These limitations are mostly attributable to the nature of the by-product gas, which is the power-generation fuel used in steel mills. A large amount of dust in the by-product gas causes sensors to malfunction, which ultimately affects the collected data. Therefore, dust must be removed from the by-product gas before a highly reliable flue gas prediction model can be developed. In addition, the model's completion level will increase if research can be conducted in a state where the calories and flow rate of the fuel are stable. Horizontal deployment to other steam generators is limited because different fuel

types are used for each type of generator. Applying the BFG-PM model requires an additional review, even if the same kind of fuel is used because each piece of equipment has different sizes, specifications, and operating conditions.

ML technology may appear to be an all-around solution for various problems in the era of the Fourth Industrial Revolution. However, the practical implementation of such cutting-edge technology often presents challenges owing to a lack of trust, causing existing operators to contemplate its adoption and hindering its enthusiastic application. The reasons for this are as follows. First, even if the equipment efficiency is slightly lower, the stable operation of power generation facilities can be maintained with a consistent combustion air supply. Consequently, there is resistance to embracing changes. Second, although using ML for operation can enhance boiler efficiency, it is essential to account for unforeseen risks. Although efficiency gains are possible, there is a need to balance them against threats that have not been experienced. Finally, introducing new technology into commercially operated power generation facilities is often met with skepticism. Engineers often hope that ML technology will be proven in terms of performance and stability in other commercial facilities before considering its application to their equipment. They were not inclined to have their equipment become a testing ground for new technology and particularly wished to avoid their equipment becoming a pilot plant for untested technology. Therefore, the verification of new technologies in different setups is preferred before their full implementation.

The following aspects should be reviewed to overcome the limitations of boiler combustion efficiency optimization. Further investment is needed in facilities for removing the dust contained in fuel to minimize the malfunction of sensors and improve the CO prediction accuracy of the BFG-PM model. The CO prediction accuracy must be improved by removing outliers of CO values through preprocessing and by selecting other variables for modeling. Additionally, the efficiency of the entire generator must be increased by optimizing the boiler combustion efficiency for facilities other than the POC.

Author Contributions: Conceptualization, S.-M.L., S.-W.C. and E.-B.L.; methodology, S.-M.L. and S.-W.C.; software, S.-M.L.; validation, S.-M.L. and S.-W.C.; formal analysis, S.-M.L. and S.-W.C.; investigation, S.-M.L. and S.-W.C.; resources, S.-M.L. and E.-B.L.; data curation, S.-M.L.; writing—original draft preparation, S.-M.L.; writing—review and editing, S.-M.L., S.-W.C. and E.-B.L.; visualization, S.-M.L. and S.-W.C.; supervision, S.-W.C. and E.-B.L.; project administration, E.-B.L.; funding acquisition, S.-M.L. and E.-B.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was sponsored by Pohang Iron & Steel Co., Ltd. (POSCO) with a grant number: POSCO Investment ID = PBS18081. The equipment used in this research is the Hyung-san No. 11 power generation boiler, the manufacturer is IHI Corporation, and the origin is Japan.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to give special thanks to Da-Som Kwon (a researcher at Pohang University of Science and Technology) for her technical support to this study. The views expressed in this paper are solely those of the authors and do not represent those of any official organization or research sponsor.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations and parameters are used in this paper:

ANN	Artificial neural network
ASME	American Society of Mechanical Engineers
BFG	Blast furnace gas
BFG-PM	Boiler Flue Gas Prediction Model

BOP	Balance of plant
CFD	Computational Fluid Dynamics
CNN	Convolutional neural network
COG	Coke oven gas
DBN	Deep belief network
ESG	Environment, Social, Governance
FOG	FINEX off gas
GA	Genetic Algorithm
GAH	Gas Air Heater
GP	Gaussian Process
IoT	Internet of things
KNN	K Nearest Neighbors
LDG	Linz-donawitz converter gas
LSFLN	Least Square Fast Learning Network
LS-SVM	Least-Squares Support Vector Machine
LSTM	Long Short-Term Memory
MES	Manufacturing Execution System
ML	Machine Learning
NSGA	Non-dominated Sorting Genetic Algorithm
PLS	Partial Least Square
PLSR	Partial Least Square Regression
PTC	Performance Test Code
SVM	Support Vector Machine
TDLS	Tunable Diode Laser Spectrometer
VIP	Variable Importance in Projection

References

1. KEPC. Korean Electricity Bill. Available online: <https://cyber.kepco.co.kr/ckepco/front/jsp/CY/E/E/CYEEHP00103.jsp> (accessed on 21 August 2023).
2. POSCO. Corporate Citizenship Report 2021. Available online: <https://www.posco.co.kr/homepage/docs/eng7/jsp/irinfo/ir-data/s91b6000030l.jsp> (accessed on 21 August 2023).
3. Ventrapati, T.; Rao, B. Efficiency and cost-benefit assessments on a typical 600MW coal-fired boiler power plant. *IJMPERD* **2019**, *9*, 201–206. <https://doi.org/10.24247/IJMPERDFEB201920>.
4. POSRI. Improving Sustainable Competitiveness in Preparation for a Circular Economy: The Case of POSCO. Available online: https://www.posri.re.kr/files/file_pdf/59/342/6926/59_342_6926_file_pdf_1531111132.pdf (accessed on 21 August 2023).
5. Babak, V.; Mokiychuk, V.; Zaporozhets, A.; Redko, O. Improving the efficiency of fuel combustion with regard to the uncertainty of measuring oxygen concentration. *East.-Eur. J. Enterp. Technol.* **2016**, *6*, 54–59. <https://doi.org/10.15587/1729-4061.2016.85408>.
6. Xiao, G.; Gao, X.; Lu, W.; Liu, X.; Asghar, A.B.; Jiang, L.; Jing, W. A physically based air proportioning methodology for optimized combustion in gas-fired boilers considering both heat release and NO_x emissions. *Appl. Energy* **2023**, *350*, 121800, <https://doi.org/10.1016/j.apenergy.2023.121800>.
7. Nemitallah, M.A.; Nabhan, M.A.; Alowaiifeer, M.; Haeruman, A.; Alzahrani, F.; Habib, M.A.; Elshafei, M.; Abouheaf, M.I.; Aliyu, M.; Alfarraj, M. Artificial intelligence for control and optimization of boilers' performance and emissions: A review. *J. Clean. Prod.* **2023**, *417*, 138109, <https://doi.org/10.1016/j.jclepro.2023.138109>.
8. Bartnicki, G.; Klimczak, M.; Ziembicki, P. Evaluation of the effects of optimization of gas boiler burner control by means of an innovative method of Fuel Input Factor. *Energy* **2023**, *263*, 125708, <https://doi.org/10.1016/j.energy.2022.125708>.
9. Santoso, H.; Ariwibowo, T.H.; Safitra, A.G. Effect of Air-Fuel Ratio to Non-premixed Burning Characteristics in Boiler Furnace Using CFD. In Proceedings of the Seminar Nasional Tahunan Teknik Mesin (SNTTM) XVI, Surabaya, Indonesia, 5–6 October 2017; pp. 92–98.
10. Pryiomov, S.; Shybetyskiy, V.; Plashykhin, S.; Kostyk, S.; Safiants, A.S.; Romanova, K.; Nizhnyk, N. Increasing the energy efficiency of cyclone dust collectors. *IJECE* **2023**, *24*, 81–96. <https://doi.org/10.1615/InterJEnerCleanEnv.2022043211>.
11. Murehwa, G.; Zimwara, D.; Tumbudzuku, W.; Mhlanga, S. Energy efficiency improvement in thermal power plants. *IJITEE* **2012**, *2*, 20–25.
12. Karri, V.S.S.K. A Theoretical Investigation of Efficiency Enhancement in Thermal Power Plants. *Mod. Mech. Eng.* **2012**, *2*, 106–113. <https://doi.org/10.4236/mme.2012.23013>.
13. Mandi, R.P.; Yaragatti, U.R. Energy efficiency improvement of auxiliary power equipment in thermal power plant through operational optimization. In Proceedings of the Power Electronics, Drives and Energy Systems (PEDES), Bengaluru, India, 16–19 December 2012; pp. 1–8.

14. Hasanuzzaman, M.; Rahim, N.A.; Hosenuzzaman, M.; Saidur, R.; Mahbubul, I.M.; Rashid, M.M. Energy savings in the combustion based process heating in industrial sector. *Renew. Sust. Energ. Rev.* **2012**, *16*, 4527–4536. <https://doi.org/10.1016/j.rser.2012.05.027>.
15. Ibrahim, T.K.; Mohammed, M.K.; Awad, O.I.; Rahman, M.M.; Najafi, G.; Basrawi, F.; Abd Alla, A.N.; Mamat, R. The optimum performance of the combined cycle power plant: A comprehensive review. *Renew. Sust. Energ. Rev.* **2017**, *79*, 459–474. <https://doi.org/10.1016/j.rser.2017.05.060>.
16. Errami, Y.; Obbadi, A.; Sahnoun, S.; Ouassaid, M.; Maaroufi, M. Performance evaluation of backstepping approach for wind power generation system-based permanent magnet synchronous generator and operating under non-ideal grid voltages. *Int. J. Power Energy Convers.* **2019**, *10*, 414–451. <https://doi.org/10.1504/IJPEC.2019.102978>.
17. Mosobi, R.W.; Gao, S. Power quality analysis of low voltage distributed generators in standalone and grid connected modes. *Int. J. Power Energy Convers.* **2021**, *12*, 267–293. <https://doi.org/10.1504/IJPEC.2021.118029>.
18. Khaleel, O.J.; Ibrahim, T.K.; Ismail, F.B.; Al-Sammarraie, A.T.; bin Abu Hassan, S.H. Modeling and analysis of optimal performance of a coal-fired power plant based on exergy evaluation. *Energy Rep.* **2022**, *8*, 2179–2199. <https://doi.org/10.1016/j.egy.2022.01.072>.
19. Zaporozhets, A. Development of software for fuel combustion control system based on frequency regulator. In Proceedings of the ICT in Education, Research, and Industrial Applications (ICTERI) 2019, Kherson, Ukraine, 12–15 June 2019; pp. 12–15.
20. Tang, Z.; Li, Y.; Kusiak, A. A deep learning model for measuring oxygen content of boiler flue gas. *IEEE Access* **2020**, *8*, 12268–12278. <https://doi.org/10.1109/ACCESS.2020.2965199>.
21. Pan, H.; Su, T.; Huang, X.; Wang, Z. LSTM-based soft sensor design for oxygen content of flue gas in coal-fired power plant. *Trans. Inst. Meas. Control* **2021**, *43*, 78–87. <https://doi.org/10.1177/01423312209323>.
22. Effendy, N.; Kurniawan, E.D.; Dwiantoro, K.; Arif, A.; Muddin, N. The prediction of the oxygen content of the flue gas in a gas-fired boiler system using neural networks and random forest. *IJ-AI* **2022**, *11*, 923–929. <https://doi.org/10.11591/ijai.v11.i3.pp923-929>.
23. Li, Z.; Li, G.; Shi, B. Prediction of Oxygen Content in Boiler Flue Gas Based on a Convolutional Neural Network. *Processes* **2023**, *11*, 990. <https://doi.org/10.3390/pr11040990>.
24. Santoso, H.M.; Nazaruddin, Y.Y.; Muchtadi, F.I. Boiler performance optimization using fuzzy logic controller. *IFAC Proc. Vol.* **2005**, *38*, 308–313. <https://doi.org/10.3182/20050703-6-CZ-1902.01780>.
25. Liu, X.; Bansal, R. Integrating multi-objective optimization with computational fluid dynamics to optimize boiler combustion process of a coal fired power plant. *Appl. Energy* **2014**, *130*, 658–669. <https://doi.org/10.1016/j.apenergy.2014.02.069>.
26. Li, G.; Niu, P.; Wang, H.; Liu, Y. Least Square Fast Learning Network for modeling the combustion efficiency of a 300WM coal-fired boiler. *Neural Netw.* **2014**, *51*, 57–66. <https://doi.org/10.1016/j.neunet.2013.12.006>.
27. Suntivarakorn, R.; Treedet, W. Improvement of boiler's efficiency using heat recovery and automatic combustion control system. *Energy Procedia* **2016**, *100*, 193–197. <https://doi.org/10.1016/j.egypro.2016.10.164>.
28. Wang, C.; Liu, Y.; Zheng, S.; Jiang, A. Optimizing combustion of coal fired boilers for reducing NOx emission using Gaussian Process. *Energy* **2018**, *153*, 149–158. <https://doi.org/10.1016/j.energy.2018.01.003>.
29. Shi, Y.; Zhong, W.; Chen, X.; Yu, A.; Li, J. Combustion optimization of ultra supercritical boiler based on artificial intelligence. *Energy* **2019**, *170*, 804–817. <https://doi.org/10.1016/j.energy.2018.12.172>.
30. Niu, Y.; Kang, J.; Li, F.; Ge, W.; Zhou, G. Case-based reasoning based on grey-relational theory for the optimization of boiler combustion systems. *ISA Trans.* **2020**, *103*, 166–176. <https://doi.org/10.1016/j.isatra.2020.03.024>.
31. Vieira, L.W.; Marques, A.D.; Schneider, P.S.; da Silva Neto, A.J.; Viana, F.A.C.; Abdel-jawad, M.; Hunt, J.D.; Siluk, J.C.M. Methodology for ranking controllable parameters to enhance operation of a steam generator with a combined Artificial Neural Network and Design of Experiments approach. *Energy AI* **2021**, *3*, 100040. <https://doi.org/10.1016/j.egyai.2020.100040>.
32. Murdoch, W.J.; Singh, C.; Kumbier, K.; Abbasi-Asl, R.; Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 22071–22080. <https://doi.org/10.1073/pnas.1900654116>.
33. Vergara, J.R.; Estévez, P.A. A review of feature selection methods based on mutual information. *Neural. Comput. Appl.* **2014**, *24*, 175–186. <https://doi.org/10.1007/s00521-013-1368-0>.
34. Chen, T.; Martin, E.; Montague, G. Robust probabilistic PCA with missing data and contribution analysis for outlier detection. *CSDA* **2009**, *53*, 3706–3716. <https://doi.org/10.1016/j.csd.2009.03.014>.
35. Ryu, S.-E.; Shin, D.-H.; Chung, K. Prediction model of dementia risk based on XGBoost using derived variable extraction and hyper parameter optimization. *IEEE Access* **2020**, *8*, 177708–177720. <https://doi.org/10.1109/ACCESS.2020.3025553>.
36. Abdulqader, D.M.; Abdulazeez, A.M.; Zeebaree, D.Q. Machine learning supervised algorithms of gene selection: A review. *Mach. Learn.* **2020**, *62*, 233–244.
37. Choi, S.W.; Lee, I.-B. Multiblock PLS-based localized process diagnosis. *J. Process Control* **2005**, *15*, 295–306. <https://doi.org/10.1016/j.procont.2004.06.010>.
38. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
39. Ding, J.G.; He, Y.H.C.; Kong, L.P.; Peng, W. Camber prediction based on fusion method with mechanism model and machine learning in plate rolling. *ISIJ Int.* **2021**, *61*, 2540–2551. <https://doi.org/10.2355/isijinternational.ISIJINT-2020-357>.

40. Zhong, F.; Liu, X.; Zhou, Q.; Hao, X.; Lu, Y.; Guo, S.; Wang, W.; Lin, D.; Chen, N. ¹H NMR spectroscopy analysis of metabolites in the kidneys provides new insight into pathophysiological mechanisms: Applications for treatment with *Cordyceps sinensis*. *Nephrol. Dial. Transplant.* **2012**, *27*, 556–565. <https://doi.org/10.1093/ndt/gfr368>.
41. Farrés, M.; Platikanov, S.; Tsakovski, S.; Tauler, R. Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *J. Chemom.* **2015**, *29*, 528–536. <https://doi.org/10.1002/cem.2736>.
42. Hastie, T.; Tibshirani, R.; Friedman, J.H.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009; Volume 2. <https://doi.org/10.1007/978-0-387-84858-7>.
43. Purseth, S.; Dansena, J.; Desai, M.S. Performance analysis and efficiency improvement of boiler—A review. *IJEAST* **2021**, *5*, 326–331. <https://doi.org/10.33564/IJEAST.2021.v05i12.057>.
44. *ASME PTC-4:2013 (Revision of ASME PTC 4-2008)*; Performance Test Code on Fired Steam Generators. ASME: New York, NY, USA, 2013.
45. Behbahaninia, A.; Ramezani, S.; Hejrandoost, M.L. A loss method for exergy auditing of steam boilers. *Energy* **2017**, *140*, 253–260. <https://doi.org/10.1016/j.energy.2017.08.090>.
46. KEPC. Electricity Supply Terms and Conditions. Available online: <https://cyber.kepco.co.kr/ckepco/front/jsp/CY/D/C/CYDCHP00401.jsp> (accessed on 21 August 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.