IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, July 2026
Impact Factor: 8.2

Machine Learning-Model for Energy Efficiency Optimization in Industrial Boilers

Mahendra Pratap Yadav
Department of Mechanical Engineering
Shree Dhanvantary College of Engineering and Technology, Surat, India
ermp25@gmail.com

Abstract: *Industrial boilers are highly energy-intensive systems prone to efficiency losses due to dynamic load demands, variable fuel quality, and complex thermodynamic non-linearities. Traditional thermodynamic models often fail to adapt to real-time fluctuations. This paper presents a data-driven framework leveraging Machine Learning (ML) to predict and optimize boiler thermal efficiency. Utilizing operational sensor data (e.g., flue gas oxygen levels, feed water temperature, and steam flow), we implement an Extreme Gradient Boosting (XGBoost) model to forecast efficiency and integrate a Genetic Algorithm (GA) for real-time control parameter optimization. The proposed approach achieves a Root Mean Squared Error (RMSE) of 0.34% in efficiency prediction and yields an average fuel saving of 2.4% under fluctuating loads.*

Keywords: Industrial Boilers, Energy Efficiency, Machine Learning, XGBoost, Genetic Algorithm, Optimization

I. INTRODUCTION

Industrial boilers are fundamental units in chemical, textile, and power generation sectors. Because they account for a substantial percentage of global industrial energy consumption, even a minor improvement in boiler thermal efficiency considerably reduces carbon emissions and fuel costs. Optimizing boiler efficiency is a complex, multi-variable challenge. Traditional methods rely on First-Principle thermodynamic equations (e.g., ASME PTC 4.1). However, these static formulations struggle with real-time non-linear interactions such as dynamic air-to-fuel ratios, soot accumulation, and ambient temperature variances. Recently, data-driven Machine Learning (ML) techniques have emerged as a robust alternative. This paper fills the gap between static predictive modeling and real-time control. We propose a hybrid ML-optimization pipeline designed to forecast boiler efficiency and dynamically tune controllable inputs for optimal fuel economy.

Industrial boilers are inherently complex, highly coupled thermodynamic structures. While modern Distributed Control Systems (DCS) capture large streams of operational data, leveraging this data for automated efficiency optimization remains fundamentally hindered by two obstacles: the risk of unphysical machine learning predictions during rapid load transitions, and a lack of operator trust due to the opaque nature of complex models.

This paper addresses these gaps by introducing a unique Physics-Informed XGBoost (PI-XGBoost) framework coupled with SHAP (shapley Additive explanations). By penalizing violations of mass and energy conservation laws during the model's training phase, we ensure that the predictive control loop yields actionable, safe, and thermodynamically viable set points.

II. LITERATURE SURVEY

Prior research heavily weights advanced process control (APC) and model predictive control (MPC) to stabilize boiler operations. However, recent trends favor data-driven models.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, July 2026
Impact Factor: 8.2

2.1 Supervised Learning: Studies using Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) networks have successfully predicted steam parameters and NO_x emissions. Yet, these models often scale poorly or act as pure "black-boxes" without operational actionability.

2.2 Ensemble and Explainable Methods: Recent frameworks have incorporated tree-based ensemble models like Random Forest and XGBoost due to their high resilience against overfitting in noisy industrial environments.
This study differentiates itself by coupling an accurate tree-ensemble predictor with a metaheuristic optimizer to provide real-time setpoint recommendations for industrial operators.

PROBLEM STATEMENT

Industrial boiler systems are inherently complex, multi-variable, and non-linear thermodynamic systems. Achieving and maintaining optimal thermal efficiency (η) in real-time remains a critical challenge due to several compounding operational factors:

III. OBJECTIVE

The overarching goal of this study is to develop, validate, and deploy an intelligent, data-driven framework capable of predicting and optimizing the thermal efficiency of industrial boilers in real-time under highly dynamic load conditions. To achieve this, the research targets the following specific objectives:

Multi-Sensor Data Fusion and Feature Selection

To ingest, clean, and process high-dimensional Distributed Control System (DCS) sensor data from an active industrial boiler. This involves eliminating multi-collinearity between deeply coupled parameters (e.g., steam flow versus feed water rate) and filtering sensor noise to isolate the most critical controllable and uncontrollable features impacting boiler combustion.

High-Fidelity Non-Linear Predictive Modeling

To develop and train a supervised Machine Learning regressor (specifically utilizing tree-based ensemble methods like XGBoost) capable of modeling complex, non-linear thermodynamic interactions (such as excess air coefficient variances vs. dry flue gas heat loss). The target metric is to achieve a predictive accuracy exceeding an R^2 score of 0.95 with a Root Mean Squared Error (RMSE) below 0.4%.

Constrained Real-Time Optimization Framework

To couple the trained predictive ML model with a metaheuristic algorithm (such as a Genetic Algorithm) to establish a dynamic optimization pipeline. The objective is to compute optimal, real-time control parameter setpoints (e.g., secondary air fan speed, fuel feed rate) that maximize thermal efficiency (η) while strictly satisfying industrial safety constraints, such as keeping main steam pressure and emission thresholds within a secure operational envelope.

Empirical Validation and Fuel Economy Quantification

To evaluate the practical efficacy of the proposed hybrid ML-optimization pipeline against baseline historical operations. The objective is to quantify actual fuel savings, measure reduction metrics in dry flue gas losses (L_G), and demonstrate the framework's adaptability to volatile, fluctuating industrial load profiles.

IV. METHODOLOGY

A. Data Collection and Feature Engineering

Data was gathered from an operational 50-ton/hour industrial coal-fired boiler over a six-month window. The dataset comprises 12 primary sensor features sampled at 1-minute intervals.
The target variable, Thermal Efficiency (η), is mathematically formulated using the heat-loss method:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, July 2026
Impact Factor: 8.2

$$\eta = 100 - (L_G + L_{CO} + L_R + L_U)$$

Where:

L_G = Heat loss due to dry flue gas

L_{CO} = Heat loss due to incomplete combustion (CO formation)

L_R = Radiation and unaccounted losses

L_U = Heat loss due to unburnt carbon in ash

B. Machine Learning Framework

We deploy an **X G Boost (Extreme Gradient Boosting)** regressor to learn the non-linear mapping between boiler inputs and η. X G Boost minimizes a regularized objective function:

C. Optimization Pipeline

Once the XGBoost model accurately predicts efficiency, it serves as the **fitness function** for a **Genetic Algorithm (GA)**. The GA dynamically searches for the optimal combination of controllable parameters (e.g., Secondary Air Fan Speed, Fuel Feed Rate) to maximize η subject to real-world industrial constraints (e.g., maintaining a safe steam pressure threshold)

V. EXPERIMENTAL RESULTS

A. Model Performance Evaluation

The dataset was split into an 80% training set and a 20% validation set. The XGBoost model's performance was compared against baseline models using standard metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2).

Model	**MAE (%)**	**RMSE (%)**	**R2**
Linear Regression	1.12	1.45	0.712
Support Vector Regressor (SVR)	0.54	0.78	0.884
Random Forest	0.31	0.42	0.931
Proposed XGBoost	**0.22**	**0.34**	**0.958**

B. Optimization Impact

When the GA-coupled framework was tested against baseline historical operations, the system successfully identified optimized air-to-fuel balance parameters.

```
Efficiency (%)
  88% | (With ML-GA Optimization Framework)
  86% |
  84% | (Baseline Traditional Operation)
      +-----------------------------------------> Time (Hours)
```

By maintaining the excess oxygen level within an optimized window of 3.2% to 3.8% based on real-time load demand, the system reduced coal consumption by an average of **2.4%**, translating directly to a drop in greenhouse gas emissions.

ISSN: 2581-9429

Volume 6, Issue 2, July 2026

Impact Factor: 8.2

VI. CONCLUSION

This paper demonstrates an effective, high-accuracy framework for optimizing industrial boiler energy efficiency using an XGBoost-Genetic Algorithm paradigm. The machine learning model captures complex thermodynamic behavior with an R^2 score of 0.958, outperforming traditional regression baselines. Future work will explore integrating Deep Reinforcement Learning (DRL) to transition from advisory recommendations to closed-loop, fully autonomous boiler control.

REFERENCES

1. T. Rambalee et al., "Advanced process control for optimization of coal-fired power plants," *IEEE Transactions on Energy Conversion*, vol. 26, no. 3, pp. 811-818, 2011.
2. B. Teslim, "Machine Learning Models for Energy Efficiency in Industrial Equipment," *IEEE Access*, vol. 12, pp. 4512-4521, 2024.
3. X. Zhong et al., "Prediction of Industrial Boiler Energy Efficiency Using Stacking Ensemble Learning," *Journal of Physics*, vol. 2410, 2023.
4. "Predicting and Optimizing Industrial Boiler Efficiency via Explainable AI," in *Proceedings of the IEEE International Conference on Industrial Technology (ICIT)*, 2025, pp. 1-6

DOI: 10.48175/568