Data in Brief 62 (2025) 111939

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

A dataset of the operating station heat rate for 806 Indian coal plant units using machine learning

Yifu Ding[a,*], Jansen Wong[b], Serena Patel[a], Guiyan Zang[a], Dharik Mallapragada[c]

[a] MIT Energy Initiative, Massachusetts Institute of Technology, United States
[b] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, United States
[c] Department of Chemical and Biomolecular Engineering, Tandon School of Engineering, New York University, United States

ARTICLE INFO

Article history:
Received 1 April 2025
Revised 11 July 2025
Accepted 23 July 2025
Available online 5 August 2025

Dataset link: A Dataset of the Operating Station Heat Rate for 806 Indian Coal Plant Units using Machine Learning (Original data)

Keywords:
Machine learning
Coal power plants
Indian power system
Station heat rate

ABSTRACT

India aims to achieve net-zero emissions by 2070 and has set an ambitious target of 500 GW of renewable power generation capacity by 2030. As of 2025, coal-based power generation accounts for approximately 47 % of the total capacity and >70 % of the total electricity generation. Upgrading and decarbonizing high-emission coal plants became a pressing energy issue. A key technical parameter in coal plant operations is the operating station heat rate (SHR), which represents the thermal efficiency during operation. Yet, the operating SHR of Indian coal plants varies and is not comprehensively documented. This study leverages existing databases to create a SHR dataset for 806 Indian coal plant units, utilizing machine learning (ML), and presents the most comprehensive coverage to date. Additionally, it incorporates environmental factors such as water stress risk and coal prices as predictive features to enhance accuracy. This dataset, accessible through our visualization platform, is designed to

* Corresponding author.
E-mail address: yifuding@mit.edu (Y. Ding).
Social media: @yifu0529,@dhariksm (Y. Ding)

https://doi.org/10.1016/j.dib.2025.111939

support energy and environmental policymaking as India advances toward its renewable energy objectives.

Specifications Table

Subject	Engineering & Materials science
Specific subject area	The operating SHR dataset of 806 Indian coal power plants as of January 2024 for power system analysis
Type of data	Tables, Raw and Processed; Figures, Processed.
Data collection	The empirical SHR data, technical specifications, and delivered coal prices of Indian coal power units are sourced from the Council on Energy, Environment, and Water (CEEW) report [1]. The locations of 806 Indian coal power units as of 2024 are obtained from the Global Energy Monitor (GEM) Global Coal Plant Tracker [2]. Additional predictive features, such as water stress risk and power system regions, are gathered and harmonized from the Water Resources Institute (WRI) [3] and Indian government reports [4].
Data source location	All data sources are based in India.
Data accessibility	Repository name: A Dataset of the Operating Station Heat Rate for 806 Indian Coal Plant Units using Machine Learning Data identification number: https://doi.org/10.5281/zenodo.13921645 Direct URL to data: https://zenodo.org/records/13921645
Related research article	Y. Ding, D. Mallapragada, R. J. Stoner, The role of coal plant retrofitting strategies in developing India's net-zero power system: A data-driven sub-national analysis, Energy for Sustainable Development, Volume 86, 2025, 101687, ISSN 0973-0826, https://doi.org/10.1016/j.esd.2025.101687.

1. Value of the Data

- A granular, plant-specific characterization of the operating SHR for the entire Indian coal plant fleet for accurate modeling and effective policy making.
- Predictive features include the technical, environmental, and spatial factors influencing SHR, such as power system regions, water stress index, and coal prices.
- Comprehensive ML algorithms accurately capture nonlinear relationships between SHR values and features.

2. Background

India is the third-largest carbon emitter globally, and its power system depends heavily on high-emission coal plants [5]. As of 2025, coal-fired power generation accounts for approximately 219 GW of capacity, representing around 47 % of the total installed capacity and generating >70 % of the country's electricity [6].

Historically, Indian coal plants have had low thermal efficiency, primarily due to the use of low-quality domestic coal and inefficient boilers. As a result, coal power generation in India produces significant carbon dioxide emissions and air pollution, contributing to many premature deaths [7]. Moreover, the majority of India's thermal power generation relies on freshwater for cooling, intensifying water stress and contributing to operational disruptions or partial shutdowns due to water shortages [8]. Between 2015 and 2018, these shortages accounted for approximately 0.2 % to 2 % of total generation losses due to outages at various thermal power plants nationwide [9]. Given the continued dependence on coal-fired plants, such challenges are expected to persist in the near to medium term. Upgrading and decarbonizing coal power generation have become India's near-term pressing energy issues.

Existing methodologies for characterizing the entire Indian coal plant fleet often oversimplify the measurement of thermal efficiency at the unit level. The SHR—the ratio of heat energy input to electricity output—is a key indicator of coal plant efficiency, helping estimate fuel consumption and power generation. A higher SHR signifies lower thermal efficiency. Without access to detailed technical parameters for each plant, such as stream pressure and temperature [10], many studies represent SHR using a single value [11,12] or a fitting curve [13] across the entire Indian coal fleet, despite significant variations due to plant designs, ambient conditions, and operating regimes. These oversimplifications can lead to underestimation of coal consumption and carbon emissions, resulting in ineffective policy decisions —a particularly pressing concern given coal's dominant role in India's energy system.

Existing databases on Indian coal plants with SHR values are incomplete or inaccurate. The most up-to-date database, the GEM coal plant tracker, records all Indian operating coal plant units, including location details, commission years, and boiler types of each coal-fired power unit [2]. However, the SHR values in this database are based on linear estimations, which may significantly deviate from the actual values. Other databases, such as those from the the Council of Energy, Environment, and Water (CEEW) [1] and the Central Electricity Authority (CEA) [14], provided the operating SHR measurements but covered only part of the Indian coal plant capacity (194 GW and 16.7 GW, respectively) and have not been updated to date. However, no database or method comprehensively covers the operating SHR for all Indian coal plants after 2023. This gap motivates the creation of a well-documented, open-access SHR database for Indian coal plants using ML prediction techniques to supplement the missing data.

3. Data Description

We present a dataset of predicted SHRs for 806 Indian coal plant units, expanding on previous works and providing the most comprehensive and up-to-date coverage. Fig. 1 shows the

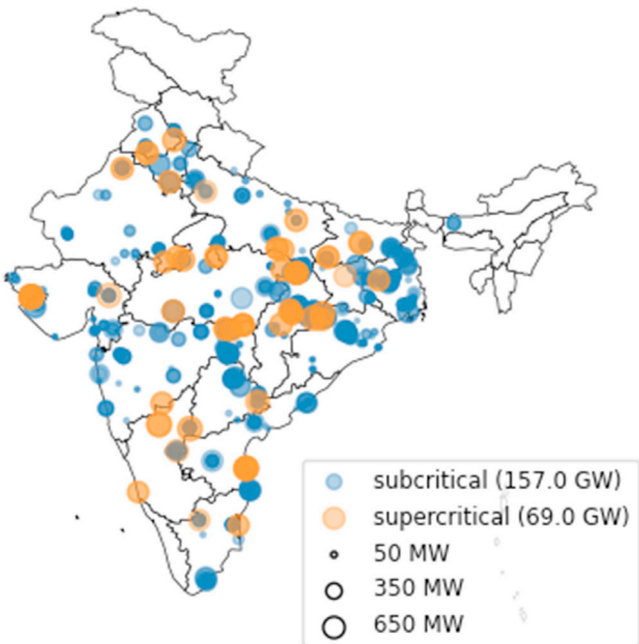

Fig. 1. Geographical locations of the 806 Indian coal plant units in the research scope [2].

Table 1
Data records of this research.

No.	Descriptions	File names
1	WRI water stress index geodata	indiawater.geojson
2	State-wise aggregated WRI water stress index	State_water_stress.csv
3	State-wise delivered coal prices	State_wise_coal_price.csv
4	30 Indian states or regions and their corresponding power system zones	30_to_5zones.csv
5	The official Indian map	india-polygon.shp
6	CEEW data with the additional features (subcritical units)	CEEW_subcritical_with_ws_price.csv
7	CEEW data with the additional features (supercritical units)	CEEW_supercritical_with_ws_price.csv
8	The original GEM data for all Indian coal plant units as of January 2024	India_coal_power_plants.csv
9	GEM data with additional predictive features	*gem*_with_ws_price.csv
10	GEM data with the predicted SHR values of all subcritical units	*gem*_predicted_subcritical.csv
11	GEM data with the predicted SHR values of all supercritical units	*gem*_predicted_supercritical.csv

geographical locations of 806 Indian operating coal plant units as of 2024, covering 226 GW in total—157 GW of 704 subcritical units and 69.2 GW of 102 supercritical units. The locations of these coal power units are sourced from the GEM Coal Plant Tracker database, released in January 2024 [2].

The datasets are available in the Zenodo repository [15] and can be downloaded on our visualization platform [16], with Table 1 detailing their contents. In this study, we predict SHR values at 50 %, 70 %, and 90 % load levels and construct detailed, piecewise heat rate curves for all coal plant units to facilitate linear computations in large-scale power dispatch and capacity expansion models. Notably, the operating SHR is a transient variable varying with time, and our model offers flexibility in predicting SHR values at any load factor by adjusting input assumptions and aligning with the specific operational regimes of individual coal plant units.

Figs. 2(a)–(d) illustrate the prediction results, with supporting code provided in Data in Brief (details on the experimental design, materials, and methods, as well as code availability). Figs. 2(a) and 2 (b) display the geographical distribution of predicted SHR values at a 50 % load factor for Indian subcritical and supercritical coal-fired power plant units. Correspondingly, Figs. 2(c) and 2(d) present the piecewise SHR curves for these units, with red line interpolations representing the heat rate approximation curves. ML predictions capture the nonlinear decline in SHR as the plant load factor increases, which shows higher thermal efficiency at higher load factors. All results are available in two files, ***gem_predicted_subcritical.csv*** and ***gem_predicted_supercritical.csv***. Table 2 summarizes the columns and content within these two files. Apart from five columns, ***bws_score***, ***coal_price***, ***Predicted_HR_50.0***, ***Predicted_HR_70.0***, and ***Predicted_HR_90.0***, the rest of the columns are based on the GEM Global Coal Plant Tracker [2].

4. Experimental Design, Materials, and Methods

Fig. 3 illustrates the four steps to predict operating SHR values of Indian coal plant units using ML models. The boiler design of coal plant units determines their operating steam pressure and temperature and, therefore, their operating SHR. Supercritical power units generally operate at higher steam pressures and temperatures, resulting in greater thermal efficiency than subcritical units [17]. Considering this effect, we separately predict coal plant units with two boiler designs, subcritical and supercritical units. Then, we extracted various predictive features and leveraged SHR measurements from the CEEW database [1] to train several ML models. Thirdly, we selected the best-performing model for each boiler type through k-fold cross-validation. Finally, we use the best-performing models to predict the SHR values for 806 coal plant units in the GEM database.

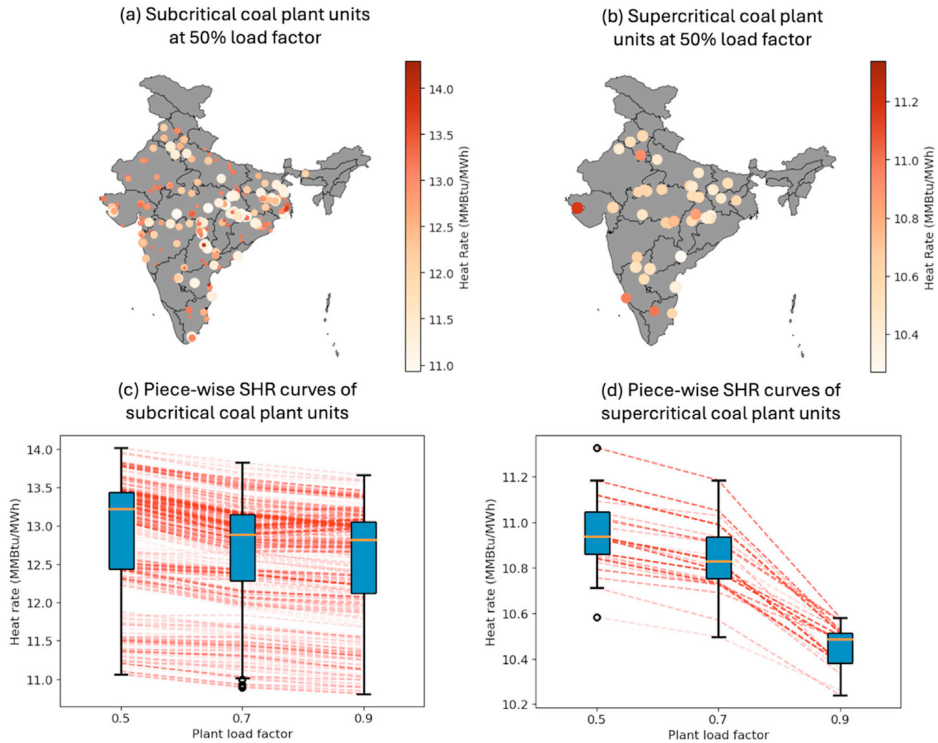

Fig. 2. Geographical distributions of the predicted SHR for Indian (a) subcritical and (b) supercritical coal-fired power units at a plant load factor of 50 %. Panels (c) and (d) show the piecewise SHR approximation curves for Indian subcritical and supercritical units at the load factor of 50 %, 70 %, and 90 %.

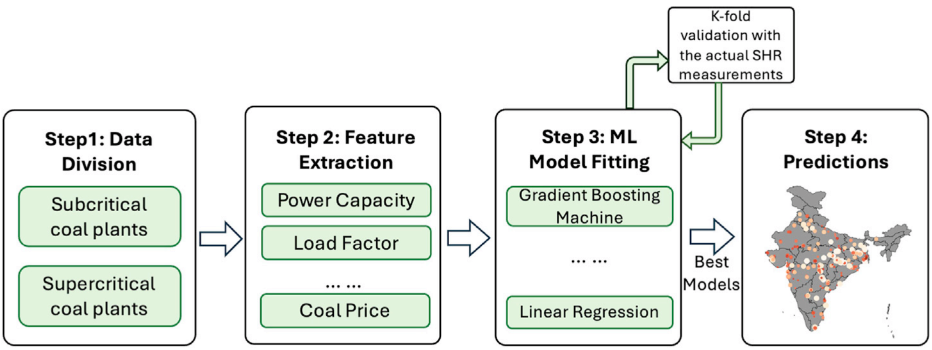

Fig. 3. Four steps to predict the operating SHR values for subcritical and supercritical coal plants using ML models.

4.1. Data preparation

We first downloaded the coal plant unit records from the GEM Global Coal Plant Tracker [2] circa January 2024. The original datasets include Indian coal plant units at various stages, including announced, mothballed, permitted, and operational units, totaling 1889. We first selected the ones currently operating, including 840 out of 1,889 units. Then, we excluded units using ultra-supercritical, circulating fluidized bed combustion, or unknown combustion technologies, as they

Table 2
Output data headers descriptor.

Columns	Units	Descriptions
Tracker ID	-	Unique tracker ID for each coal plant unit in the GEM database
Wiki Page	-	Wiki page address of each coal plant unit in the GEM database
Country	-	Country of coal plant units (India)
Subnational unit (province, state)	-	Subnational regions (province, state) of coal plant units
Unit	-	Unit number in the coal power plant
Plant	-	Name of coal power plant
Capacity	MW	Nameplate power capacity of coal plant units
Year	-	Year of commissioning for coal plant units
Combustion technology	-	Combustion technology employed in each coal plant unit
Latitude	-	Latitude of the unit location
Longitude	-	Longitude of the unit location
Heat rate	Btu/kWh	The original calculated SHR values of coal plant units in the GEM database
Remaining plant lifetime	years	The remaining lifetime of each unit given a 50-year lifetime
bws_score	-	WRI water stress index at the location of coal plant units
coal_price	$/MMBtu	Delivered coal prices of coal power units in 30 Indian regions
Predicted_HR_50.0	MMBtu/MWh	Predicted SHR values of subcritical or supercritical units at the load factor of 50 %
Predicted_HR_70.0	MMBtu/MWh	Predicted SHR values of subcritical or supercritical units at the load factor of 70 %
Predicted_HR_90.0	MMBtu/MWh	Predicted SHR values of subcritical or supercritical units at the load factor of 90 %

Table 3
Comparison of training datasets and predicted targets.

	Training datasets	Prediction targets
Data sources	CEEW [1]	GEM Coal Plant Tracker [2]
Coal plant unit coverage	541 Coal Plant Units (194 GW)	840 Coal Plant Units (234 GW)
Date of data measured or created	2017–2020	2024
Resolution of technical specifications	Unit-level boiler type, age as of 2020, power capacity and average load factor over 30 months	Unit-level boiler type, age as of 2024, power capacity
Resolution of location details	Subnational level (province, state)	Unit level (latitude, longitude)

accounted for <5 % of the total capacity. The remaining 806 coal plant units are the target for SHR predictions.

The training dataset is based on the CEEW report [1], which covers 541 coal plant units and contributed to 98 % of the total coal plant capacity in 2020 [18]. The SHR measurements of these plants span from September 2017 to February 2020, covering 30 months before the COVID-19 pandemic. It also records the technical details of each coal plant unit, including the boiler design, state, age as of 2020, power capacity, and average load factors across 30 months. Table 3 illustrates differences between the training datasets and prediction targets.

4.2. Feature extraction

We selected a range of predictive features covering technical parameters, environmental factors, and geographical locations. Specifically, we included six features: power capacity, plant age, load factor, water stress index, delivered coal price, and power system regions. While other me-

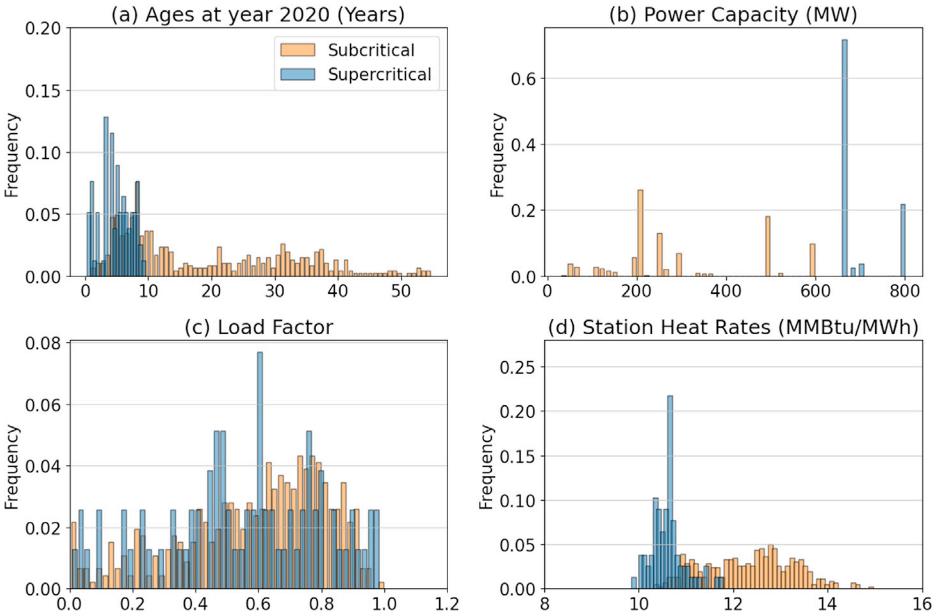

Fig. 4. Distributions of the unit-level characteristics in the CEEW database [1]: (a) Plant ages at year 2020 (years) (b) Power capacity (MW) (c) Average load factor, and (d) Station heat rate (MMBtu/MWh); The orange bars represent feature distributions of subcritical coal plant units, and the blue bars represent feature distributions of supercritical coal plant units.

teorological variables, such as the maximum daily air temperature [19], were considered in the correlation analysis, they were excluded from this study due to their substantially lower correlation with the predicted target compared to water stress. All features are treated as continuous variables, except for the power system regions. We converted the five different power system regions into five distinct binary features using one-hot encoding, indicating whether a coal power plant is in each specific region. Before predictions, we scaled all predictor variables to values between 0 and 1 using the Min-Max scaling method, ensuring consistent variable ranges.

The technical features include the age as of 2020, average load factor, and power capacity of each coal unit. Fig. 4 shows these feature distributions with SHR measurements from the CEEW report [1]. The plant load factor in the training data is the average ratio of the actual power generated to the potential power generation based on the installed capacity over the plant's operating period, and its value is between 0 and 1 (Fig. 4(c)). The feature distributions of subcritical and supercritical units exhibit distinct differences. The median age of supercritical units is around five years, much less than that of subcritical units (Fig. 4(a)). The SHR values for subcritical units exhibit a relatively wider distribution, with many power plants displaying very high heat rates (Fig. 4(d)).

The environmental factors include the location-based water stress index and the delivered coal price. According to the WRI, the raw value of water stress is calculated from the ratio of water demand to the available freshwater supply, which includes agricultural, domestic, industrial, and power generation usage [3]. The water stress index is classified into five levels based on the water stress raw value ranges, as listed in Table 4. For example, an extreme water stress level above 0.8 means the water withdrawal exceeds 80 % of the available water supply. Under high or extremely high water stress, the government could restrict water use during dry seasons or require alternative cooling methods, such as dry cooling, which affects the thermal efficiency of coal plant units [3]. As shown in Fig. 5(a), a significant part of India is classified as under "high" or "extremely high" water stress levels. The water stress index is harmo-

Table 4
WRI water stress index and classifications [3].

Water stress level	Water stress index	Raw value
Low	<1	<10 %
Low to medium	1–2	10–20 %
Medium to high	2–3	20–40 %
High	3–4	40–80 %
Extremely high	>4	>80 %

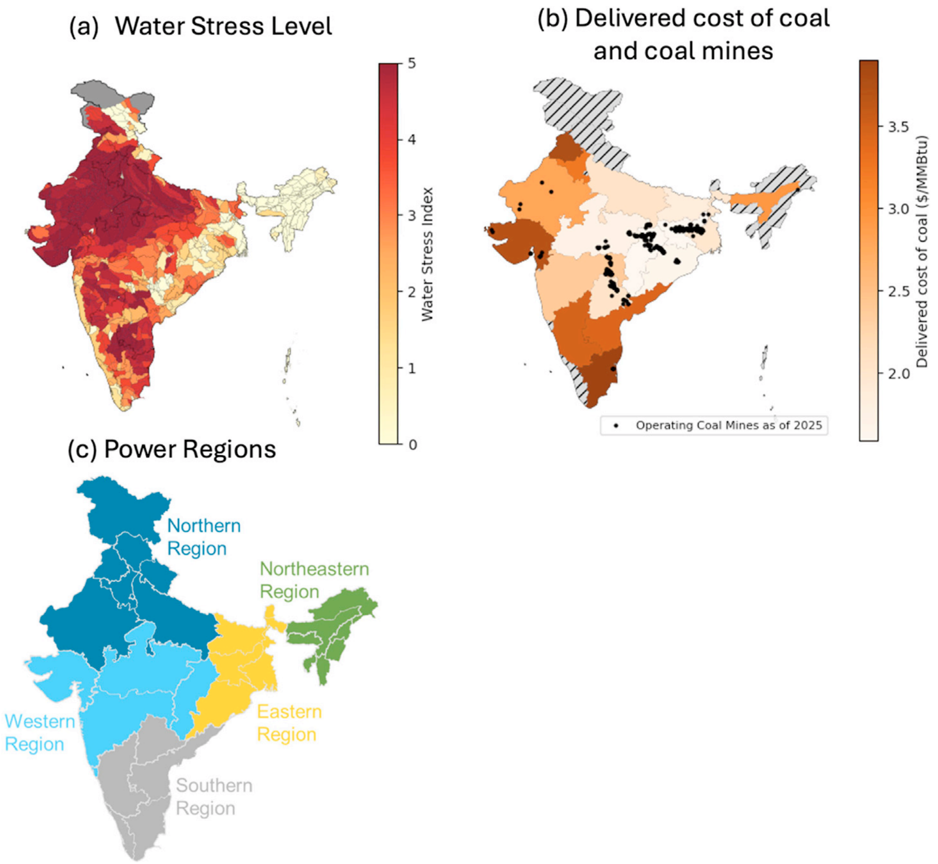

Fig. 5. Environmental and geographical predictive features: (a) Water stress level [3], (b) Delivered coal prices [1], Locations of operating Indian coal mines as of 2025 [20], and (c) Power system regions [4]. The grey or shaded areas in the maps (a) and (b) mean that no coal plant data is available, or no coal plant has been built in the area.

nized with the state and plant levels for the training dataset and prediction targets, respectively (***State_wise_coal_price.csv; gem_with_ws_price.csv*** in Table 1).

The state-wise delivered coal prices are extracted from the CEEW report [1], ranging from $1.59 to $3.90/MMBtu (Fig. 5(b)). The coal price incorporates the production (i.e., domestic or imported) and transportation costs of coal (i.e., delivered coal price). The eastern regions of India have significantly lower coal prices than the rest of the country, as most coal mines are in these regions to supply nearby coal power generation, thereby incurring lower coal transportation costs.

The geographical features of Indian power regions are considered predictive features. The Indian power grid is divided into five areas - Northern, Eastern, Western, North Eastern, and Southern, as shown in Fig. 5(c) [4]. Within each region, thermal power plants are often dispatched based on heuristic rules, which partially impact the operation and efficiency of coal plant units based on their locations [13].

4.3. ML prediction model selection and fitting

We built a set of ML models to predict the operating SHR of subcritical and supercritical units separately. There are six ML models, including gradient boosting machine (GBM), XGBoost (XGBM), random forest regressor (RF), decision tree regressor (DT), support vector regression (SVR), and k-Neighbors regressor, and two statistical regression models for comparison, including ridge regression and linear regression. These models were chosen due to their diverse underlying algorithms, which comprehensively explore potential solutions.

The training dataset was divided into the training and test parts, including SHR measurements for 541 coal plant units from the CEEW report [1] and three prediction features. The training set comprises 432 of 541 plant units (i.e., 370 subcritical and 62 supercritical units). The remaining 109 plant units (i.e., 93 subcritical and 16 supercritical units) were allocated to the test part for model validation.

To set up robust ML models, we employed the grid search to tune six ML models. Grid search optimizes the hyperparameters for each model through five-fold cross-validation and then selects the optimal combination that minimizes mean squared error (MSE) across all splits. We rotated the training and test datasets within each fold to ensure that all 541 coal units were comprehensively tested. At least two key hyperparameters were chosen for each ML model. The search ranges for these hyperparameters were carefully defined through multiple experiments to ensure the optimal value falls within the range without reaching the boundaries. The selected hyperparameters and search details are documented and explained in the Code availability subsection. In summary, each model underwent 371 hyperparameter search combinations, leading to 1855 model fits per unit type.

Model performances were evaluated using an R^2 score and three prediction accuracy metrics: MSE, mean absolute error (MAE), and mean absolute percentage error (MAPE). The R^2 score presents the goodness-of-fit of different prediction models. The values typically range from 0 to 1. The R^2 score of 1 means a perfectly fitted model, and the R^2 score of 0 or below indicates no added value for using the model compared to averaging the dataset. Tables 5 and 6 display the metric values across eight models for subcritical and supercritical units, with the best values highlighted in bold. Based on the R^2 score and three prediction accuracy metrics, the GBM and k-nearest neighbors regressors are the best models for subcritical and supercritical units, respectively. Fig. 6 shows the parity plots of the actual and predicted SHR values for both unit types using the two best ML models.

Table 5
Three prediction accuracy metrics and R^2 scores for eight models (subcritical units).

Models	MSE	MAE	MAPE	R^2
Gradient Boosting Machine	**0.0295**	**0.0995**	**0.0081**	**0.9664**
Random Forest Regressor	0.0432	0.1280	0.0104	0.9507
Decision Tree Regressor	0.0632	0.1571	0.0127	0.9283
XGBoost	0.0327	0.1141	0.0093	0.9625
Linear Regression	0.0368	0.1176	0.0095	0.9581
Ridge Regression	0.0378	0.1236	0.0100	0.9569
SVR	0.0437	0.1211	0.0097	0.9503
KNeighbors Regressor	0.0639	0.1493	0.0121	0.9264

Table 6
Three prediction accuracy metrics and R^2 scores for eight models (supercritical units).

Models	MSE	MAE	MAPE	R^2
Gradient Boosting Machine	0.0465	0.1270	0.0118	0.6455
Random Forest Regressor	0.0629	0.1721	0.0161	0.4584
Decision Tree Regressor	0.0877	0.1911	0.0179	0.3100
XGBoost	0.0478	0.1315	0.0124	0.5916
Linear Regression	0.0529	0.1869	0.0176	0.5044
Ridge Regression	0.0997	0.2319	0.0217	0.0924
SVR	0.0338	0.1308	0.0122	0.7134
KNeighbors Regressor	**0.0222**	**0.0988**	**0.0093**	**0.8105**

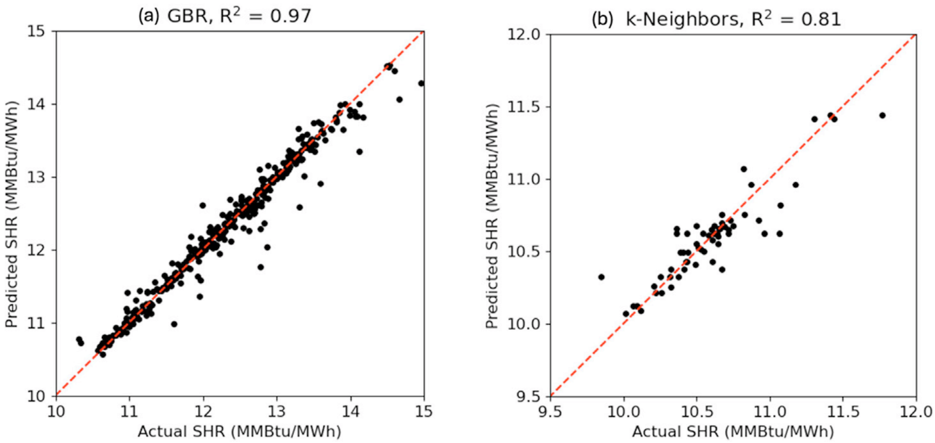

Fig. 6. R^2 scores and the parity plots for the two best ML models: (a) GBR for subcritical units; (d) k-neighbors regression for supercritical units.

4.4. Feature importance analysis and operational implications

Finally, we assessed the importance of features to find the key factors influencing the predicted SHR. We calculated the Shapley Additive Explanations (SHAP) value and ranked the importance of various features based on the absolute mean of the SHAP value. As shown in Figs. 7(a) and (b), the three most important features for subcritical units are power capacity, age, and plant load factor. This finding aligns with the wide range of power capacity and age

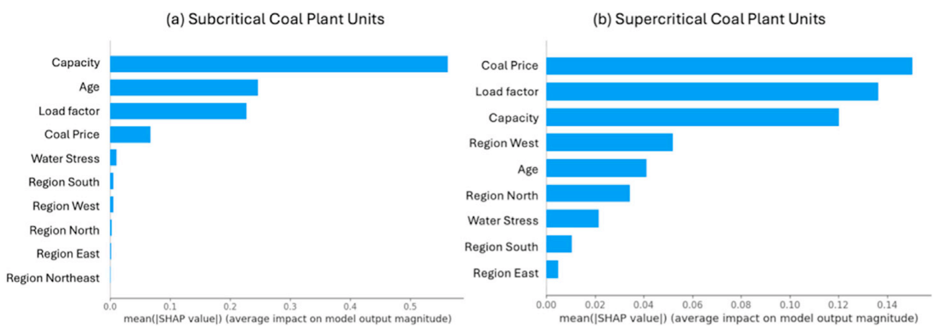

Fig. 7. The absolute mean SHAP values for SHR predictive features for (a) subcritical units and (b) supercritical units.

Table 7
Code to generate the database.

No.	Descriptions	File names
1	Data harmonization for water stress index and coal price features	data_harmonization.ipynb
2	Grid search for hyperparameter tuning of ML prediction models (subcritical units)	GRID_search_subcrit.ipynb
3	Grid search for hyperparameter tuning of ML prediction models (supercritical units)	GRID_search_supercrit.ipynb
4	Prediction framework for subcritical units based on the best-performing ML model	run_models_subcritical.ipynb
5	Prediction framework for supercritical units based on the best-performing ML model	run_models_supercritical.ipynb
6	Python environmental dependencies of the code	enviromental.yml

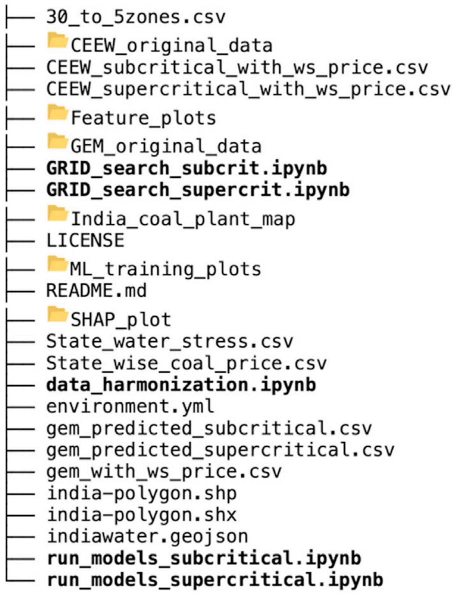

Fig. 8. All files in the GitHub repository [22].

of subcritical coal plants in India. The coal price is the most important feature for supercritical units, followed by load factor and power capacity. Our results show that plant age does not significantly influence the SHR of supercritical units, primarily due to the relatively young age profile of supercritical coal plants, which are typically <10 years old. Remarkably, the water stress has a limited impact on the SHR predictions for both unit types. This indicates that most Indian coal plants operate independently of local water stress conditions, although their operations are influenced by water shortage.

Based on the SHR prediction results and feature importance analysis, and considering India's continued dependence on coal in the near to medium term, it is essential to adopt coal plant operation strategies that emphasize improved water management and overall thermal efficiency. Potential measures include utilizing recycled wastewater [9], enhancing the operational flexibility of supercritical coal units [1], and phasing out the reliance on small, aging supercritical plants [21].

4.5. Code availability

All datasets, code, and environmental dependencies are implemented in Python and can be accessed in the following GitHub repository [22]. We predict the up-to-date number of coal plant units in India (i.e., 806 units as of January 2024) and built a visualization platform for downloading the dataset [16]. The code used to generate features and predict the entire SHR database is available in the same GitHub repository [22]. The detailed code descriptions are in Table 7. The file structure in the GitHub repository is presented in Fig. 8, with five Python files used to create the database highlighted in **bold**. All the plots presented in this paper can be found in the ***India_coal_plant_map*** and ***Feature_plots*** folders, along with the corresponding code. The original datasets for the training set and the prediction targets are in the ***GEM_original_data*** and ***CEEW_original_data*** folders, respectively.

Limitations

Several key considerations arise when applying new datasets for coal plants or incorporating additional prediction features, following the four steps outlined in Fig. 3. First, our training dataset does not account for novel combustion technologies (e.g., ultra-supercritical), since these units contribute <8 % of the current total coal plant power capacity and lack recorded SHR measurements in the existing literature. Second, the grid search process took approximately 35 min for subcritical and supercritical coal plant units. When introducing new ML models or updating datasets, adjustments to the number of searched hyperparameters and their ranges may be necessary, which can impact the overall computation time. Third, access to detailed technical parameters—such as transient steam pressure and temperature—would enable more accurate modeling of SHR. However, such datasets are not currently available for many coal plants in India.

Data Availability

A Dataset of the Operating Station Heat Rate for 806 Indian Coal Plant Units using Machine Learning (Original data) (Zenodo).

CRediT Author Statement

Yifu Ding: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Formal analysis, Data curation; **Jansen Wong:** Visualization, Validation, Software, Resources, Methodology, Formal analysis, Data curation; **Serena Patel:** Visualization, Validation, Methodology, Formal analysis, Data curation; **Guiyan Zang:** Supervision, Validation, Project administration; **Dharik Mallapragada:** Writing – review & editing, Supervision, Project administration, Methodology, Validation, Data curation.

Acknowledgements

Yifu Ding was supported by IHI Corporation, Japan. Serena Patel was supported by an MIT TATA fellowship. Jansen Wong was supported by the MIT Undergraduate Research Opportunities Program (UROP). This work recognizes the recipient of the 2024 MIT Open Data Prize.

Declaration of Competing Interest

The authors declare no competing interests.

References

[1] K. Ganesan, D. Narayanaswamy, Coal Power's trilemma: variable cost, efficiency and financial solvency, CEEW (Jul. 2021) Accessed: Feb. 28, 2023. [Online]. Available: https://www.ceew.in/sites/default/files/CEEW-study-on-thermal-decommissioning-coal-electricity-power-plants.pdf .
[2] Global Energy Monitor, "Global Coal Plant Tracker," Global Coal Plant Tracker January 2024 release (2024), Accessed: Feb. 28, 2024. [Online]. Available: https://globalenergymonitor.org/projects/global-coal-plant-tracker/
[3] R. Hofste, et al., Aqueduct 3.0: updated decision-relevant global water risk indicators, WRIPUB (2019), doi:10.46830/writn.18.00146.
[4] U.S. EIA, Interregional Electricity trade in India, Independent Statistics &Analysis (2020), Accessed: Mar. 19, 2024. [Online]. Available: https://www.eia.gov/outlooks/ieo/pdf/IEO2020_IIF_India.pdf
[5] IEA, "Country profile: India." (2022), Accessed: May 08, 2024. [Online]. Available: https://www.iea.org/countries/india/emissions
[6] NITI Aayog, "India Climate & Energy dashboard", 2025, Accessed: Jun. 18, 2025. [Online]. Available: https://iced.niti.gov.in/energy/electricity/generation/capacity
[7] M. Cropper, et al., The mortality impacts of current and planned coal-fired power plants in India, Proc. Natl. Acad. Sci. U.S.A. 118 (5) (Feb. 2021) e2017936118, doi:10.1073/pnas.2017936118.
[8] T. Luo, D. Krishnan, and S. Sen, "Parched power: water demands, risks, and opportunities for India's power sector," 2018. Accessed: Mar. 19, 2024. [Online]. Available: https://files.wri.org/d8/s3fs-public/parched-power-india-0130.pdf
[9] S.S. Das, R. Majumdar, A.V. Krishnan, Exploring urban water-energy nexus: a case study of thermal power plants in Raichur and Ballari districts in Karnataka, Nat. Resour. Forum (2024) 1–31, doi:10.1111/1477-8947.12507.
[10] W. Muhammad Ashraf, V. Dua, Driving towards net-zero from the energy sector: leveraging machine intelligence for robust optimization of coal and combined cycle gas power stations, Energy Convers. Manage. 314 (Aug. 2024) 118645, doi:10.1016/j.enconman.2024.118645.
[11] R. Deshmukh, A. Phadke, D.S. Callaway, Least-cost targets and avoided fossil fuel capacity in India's pursuit of renewable energy, Proc. Natl. Acad. Sci. U.S.A. 118 (13) (Mar. 2021) e2008128118, doi:10.1073/pnas.2008128118.
[12] Central Electricity Authority, "Draft report on optimal generation capacity mix for 2029-2030", (April 2023), Accessed: Mar. 07, 2024. [Online]. Available: https://cea.nic.in/wp-content/uploads/irp/2023/05/Optimal_mix_report__2029_30_Version_2.0__For_Uploading.pdf.
[13] S. Sengupta, et al., Subnational implications from climate and air pollution policies in India's electricity sector, Science 378 (6620) (Nov. 2022) eabh1484, doi:10.1126/science.abh1484.
[14] Central Electricity Authority, "Mapping of 85 pulverized coal fired thermal power generating units in different states (Under Indo-German Energy Program (IGEN))." (2010), Accessed: Mar. 17, 2024. [Online]. Available: https://cea.nic.in/wp-content/uploads/2020/04/report_85_pul_coal.pdf.
[15] Y. Ding, J. Wong, S. Patel, D. Mallapragada, G. Zang, and R. Stoner, "A dataset of the operating station heat rate for 806 indian coal plant units using machine learning." zenodo, Oct. 11, 2024. doi: https://doi.org/10.5281/zenodo.13921645.
[16] J. Wong, Y. Ding, and S. Patel, "Visualization platform for datasets of the operating station heat rate for 806 indian coal plant units using machine learning", (2024), [Online]. Available: https://states-mode–admirable-creponne-5225d2.netlify.app/.
[17] U.S. aid, "Best practices manual for Indian supercritical plants: an abstract." Nov 2013, Accessed: Sep. 21, 2024. [Online]. Available: https://sarepenergy.net/wp-content/uploads/2022/05/Best-Practices-Manual.pdf.
[18] Central Electricity Authority, Government of India, "Installed capacity Report (January 2020)." 2020, Accessed: Mar. 20, 2025. [Online]. Available: https://cea.nic.in/installed-capacity-report/?lang=en.
[19] F.F. Vieira, et al., Statistical analysis of extreme temperatures in India in the period 1951–2020, Theor. Appl. Climatol. 152 (1–2) (Apr. 2023) 473–520, doi:10.1007/s00704-023-04377-5.
[20] Global Energy Monitor, "Global Coal Mine Tracker." May 2025. Accessed: Jun. 19, 2025. [Online]. Available: https://globalenergymonitor.org/projects/global-coal-mine-tracker/.
[21] Y. Ding, D. Mallapragada, R.J. Stoner, The role of coal plant retrofitting strategies in developing India's net-zero power system: a data-driven sub-national analysis, Energy Sustain. Develop. 86 (Jun. 2025) 101687, doi:10.1016/j.esd.2025.101687.
[22] Y. Ding and J. Wong, "Code for India SHR prediction." (2024), [Online]. Available: https://github.com/yifueve/India_SHR_dataset.