Arabian Journal for Science and Engineering
https://doi.org/10.1007/s13369-026-11143-y

RESEARCH ARTICLE-ELECTRICAL ENGINEERING

Time-Aware Machine Learning for Biomass Power Output Estimation Using SCADA Data

Ezgi Güney[1] · **Memnun Demir**[2]

Received: 19 October 2025 / Accepted: 3 February 2026

Abstract

Accurate short-term estimation of electrical power output in biomass power plants remains challenging due to the nonlinear and dynamically coupled nature of thermochemical conversion processes, fuel heterogeneity, and pronounced thermal inertia. Conventional physics-based models, while effective for steady-state analysis, often fail to capture the high-frequency dynamics required for real-time monitoring and decision-support applications. This study proposes a data-driven framework for short-term power output estimation using high-resolution Supervisory Control and Data Acquisition (SCADA) data collected from an operational industrial biomass power plant. A large-scale SCADA dataset comprising several hundred thousand time-stamped records is used to model the relationship between seven key thermodynamic and operational variables and net electrical power output. Multi-layer perceptron (MLP), random forest (RF), gradient boosting regressor (GBR), and support vector regression (SVR) are evaluated under two distinct validation strategies: (i) a conventional random train–test split and (ii) a temporally blocked cross-validation scheme preserving causal order. Under random sampling, RF attains the highest apparent accuracy (R^2 = 0.9687), whereas MLP exhibits lower performance (R^2 = 0.8492), highlighting sensitivity to instantaneous regression assumptions. When temporal continuity is enforced, predictive performance improves consistently across all models. In the blocked validation stage, GBR and RF achieve R^2 values of 0.9983 and 0.9973, respectively, while MLP demonstrates a substantial performance increase (R^2 = 0.9865). Time-domain analysis further reveals that ensemble-based models provide smoother tracking of short-term fluctuations, whereas temporally aligned evaluation significantly improves the physical consistency of neural network predictions. These results demonstrate that temporally consistent validation is essential for reliable SCADA-based modeling of biomass power generation and provide a practical foundation for real-time monitoring and decision-support applications in industrial biomass power plants.

Keywords Biomass · SCADA · Machine learning · Temporal information leakage · Short-term power forecasting

1 Introduction

The global transition toward sustainable, low-carbon, and resource-efficient energy systems has significantly increased the strategic importance of renewable power generation technologies. In this context, biomass-based power plants occupy a distinctive position by enabling the conversion of agricultural residues, forestry waste, and organic by-products into dispatchable electrical energy, thereby supporting circular economy principles while contributing to grid stability and energy security [1–3]. Unlike intermittent renewable sources such as wind and solar, biomass power plants can provide controllable and continuous generation, making them particularly valuable for balancing power systems with high renewable penetration [4].

Despite these advantages, biomass power generation systems are characterized by substantial operational complexity. Electricity production relies on multistage thermochemical conversion processes involving fuel preparation, combustion, steam generation, and turbine-driven power conversion. These stages are strongly coupled and governed by pro-

✉ Ezgi Güney
eguney@sinop.edu.tr

Memnun Demir
mdemir@sinop.edu.tr

[1] Electrical and Energy, Sinop University, Selanik, Sinop 100190, Sinop, Turkey

[2] Computer Technologies, Sinop Üniversitesi, Atatürk, Sinop 57600, Gerze, Turkey

Published online: 19 February 2026

nounced thermal inertia, nonlinear interactions, and delayed system responses [5–7]. Variations in key process variables, such as steam flow rate, boiler pressure, and combustion conditions, therefore do not instantaneously translate into changes in electrical output. Instead, their effects propagate through the system over finite time intervals, resulting in temporally dependent and dynamically evolving behavior. Fuel heterogeneity, including variations in moisture content, ash composition, and calorific value, further amplifies this complexity by introducing stochastic disturbances that affect combustion stability and downstream conversion efficiency [8].

Traditional physics-based models have been extensively used to describe biomass power plants, particularly for steady-state analysis and design optimization [9]. However, such models often require detailed system knowledge and are computationally demanding, which limits their applicability for high-frequency monitoring, fast operational diagnostics, and real-time decision-support applications. In recent years, machine learning (ML) and deep learning (DL) techniques have emerged as powerful alternatives for data-driven modeling of complex energy systems. While solar photovoltaic and wind power forecasting have achieved rapid performance gains due to relatively direct relationships between environmental drivers and power output [10–13], biomass power plant modeling remains more challenging. This difficulty arises primarily from delayed causality, feedback-controlled dynamics, and strong process coupling inherent to boiler–steam–turbine subsystems [14, 15].

Despite the growing body of ML- and DL-based studies in biomass power plant applications, short-term estimation of net electrical power output under realistic industrial operating conditions remains insufficiently addressed. Existing research has predominantly focused on intermediate process variables, such as combustion-related indicators, virtual sensing of flue-gas characteristics, vision-based combustion state assessment, or upstream fuel property prediction, rather than directly modeling the temporally evolving power conversion chain. Moreover, a substantial portion of the literature continues to rely on regression formulations and evaluation strategies that implicitly assume instantaneous input–output relationships and independent, identically distributed samples. When applied to high-frequency Supervisory Control and Data Acquisition (SCADA) data, such practices frequently involve random train–test splitting, which violates temporal causality and leads to systematic information leakage. As a result, reported performance metrics may substantially overestimate true predictive capability and fail to reflect real-time deployment conditions.

Consequently, a clear research gap exists at the intersection of short-term net power output modeling, temporal dependency-aware feature representation, and causally consistent validation strategies using real-field industrial SCADA data. To address these gaps, this study makes the following key contributions:

- A systematic comparison of random sampling and temporally blocked validation strategies for biomass power output estimation using high-frequency SCADA data.
- A quantitative and time-domain demonstration of the impact of temporal information leakage on apparent model performance.
- An evaluation of lag-aware ML models that incorporate delayed system responses without altering core learning architectures.
- The provision of a physically consistent and practically applicable modeling framework suitable for real-time monitoring, anomaly detection, and decision-support applications in industrial biomass power plants.

The remainder of the paper is organized as follows. Section 2 describes the case study, data acquisition, pre-processing, and modeling methodology. Section 3 presents and discusses the results obtained under both evaluation stages. Finally, Sect. 4 concludes the paper and outlines directions for future research.

1.1 Related Work on ML/DL-Based Modeling of Biomass Power Plants

From a ML/DL perspective, the biomass literature has progressed along several complementary directions, yet with a notable emphasis away from short-term net electrical power estimation. A prominent research thread focuses on soft sensing and virtual sensing of combustion-related variables that are difficult to measure reliably or exhibit significant measurement lag, such as flue gas oxygen content and other combustion quality indicators. Recent studies demonstrate that DL architectures, particularly LSTM-based predictors and hybrid optimization frameworks, can act as surrogate sensors to provide more timely feedback for combustion control and efficiency enhancement [16–18]. While these approaches confirm the strongly nonlinear and history-dependent nature of boiler combustion dynamics, they predominantly target intermediate process variables rather than the net electrical power output of the plant.

Another important direction involves the enrichment of conventional measurements with high information content signals, most notably through vision-based DL. Image-based neural network models have been employed to complement standard operational measurements with real-time flame imaging in order to capture fast combustion dynamics that are not fully reflected in SCADA data [19]. More recent work extends this concept to the characterization of combustion regimes and thermal power states, including biomass co-combustion scenarios, using convolutional neural network (CNN)-based pipelines [20]. Although these studies highlight the potential of DL to capture rapid combustion phenomena, their primary modeling targets typically remain thermal power or combustion state rather than short-term net electrical power delivered by the plant.

A third major body of research concentrates on fuel-centric prediction tasks, such as estimating the Higher Heating Value (HHV) of biomass fuels from proximate and ultimate analyses. Comparative studies report strong performance of ensemble-based and kernel-based ML models for HHV prediction, supporting their use in upstream fuel characterization and supply-chain planning contexts [21, 22]. However, despite their relevance for fuel assessment, these approaches do not directly address the operational, time-dependent power conversion chain in industrial biomass power plants.

Despite this expanding body of ML/DL-based research, short-term power output estimation in industrial biomass power plants remains subject to two recurring limitations. First, many studies continue to rely on regression formulations that implicitly assume instantaneous input–output relationships and independent, identically distributed samples, even when high-frequency SCADA data are available [23, 24]. Second, and more critically for real-world deployment, a substantial portion of the literature evaluates model performance using random train–test splits. In time-ordered datasets, such practices violate temporal causality and can induce systematic information leakage, leading to overly optimistic performance estimates that are not reproducible under real operating conditions [25, 26]. This issue is well documented in the time-series evaluation literature, where blocked or otherwise time-respecting cross-validation strategies are recommended to preserve causal order and obtain robust generalization error estimates [27, 28]. Beyond classical time-series settings, data leakage has also been shown to materially distort performance claims in applied modeling studies when splitting strategies mishandle dependence structures or repeated measurements [29].

2 Methodology

2.1 Case Study Description

This study was conducted at a medium-scale biomass power plant with an installed capacity of approximately 10 MW, located in the Sinop Organized Industrial Zone, Turkey. The facility operates under real industrial conditions and provides a representative case for evaluating data-driven power output estimation using high-frequency operational data.

The plant primarily utilizes forest residues, including wood chips, bark, and root biomass, together with agricultural by-products originating from the wood-processing industry. Prior to combustion, the biomass feedstock undergoes size reduction, drying, and, when required, homogenization to ensure stable and controlled combustion behavior. The plant operates according to a conventional Rankine cycle and consists of four main subsystems: the fuel handling system, the water–steam system, the gas system, and the steam turbine system. High-pressure steam, typically generated at approximately 61 bar and 500–600 °C, is expanded in the steam turbine to convert thermal energy into mechanical power and subsequently into electrical energy. All major process variables and electrical output parameters are continuously monitored and recorded through a SCADA system.

2.2 Data Acquisition and Variable Set

Operational data were obtained from the SCADA system of the biomass power plant. The dataset consists of 337,856 time-stamped records collected from plant-wide sensors monitoring key thermodynamic and operational parameters. This high-resolution dataset enables statistically robust representation of plant behavior and supports detailed analysis of short-term dynamics under real operating conditions. The dataset includes seven input variables and one output variable corresponding to the net electrical power output. The units, physical meanings, and operational relevance of all variables are summarized in Table 1.

These parameters represent critical process indicators that directly or indirectly influence power generation performance in biomass-based energy systems. All variables were synchronized using their time stamps to preserve temporal continuity. Missing values and anomalous measurements were identified and handled during the pre-processing stage to ensure data integrity prior to model development.

Fuel-related properties, such as moisture content and chemical composition, were not included as direct model inputs due to the absence of real-time SCADA measurements. In practical plant operation, these properties are indirectly regulated through operator-controlled fuel blending strategies to maintain stable combustion conditions. Within the considered operational window, their short-term impact on electrical power output is therefore assumed to be limited. The multivariate SCADA time-series structure employed in this study is illustrated in Fig. 1, which presents the temporal evolution of the seven process variables together with the electrical output power.

During the observation period, the plant operated predominantly under stable production conditions, with the output power maintained within a narrow range of approximately 8–9 MW. Superimposed short-term fluctuations reflect routine operational adjustments and inherent combustion dynamics. The process variables exhibit heterogeneous dynamic behavior. Boiler pressure and temperature remain within relatively narrow bounds due to regulatory control,

Table 1 Description of process parameters, units, and operational explanations used in the study

Parameter	Unit	Explanation
Steam flow rate	Ton/h	Primary load indicator
Boiler temperature	°C	Thermal efficiency constraint
Furnace internal vacuum	Bar	Ensures stable combustion and prevents gas leakage
Boiler pressure	Bar	Turbine inlet condition
Drum level	mmH_2O	Boiler safety and stability
Deaerator level	%	Feedwater conditioning
Flue gas O_2	% O_2	Combustion efficiency
Output power	MW	Target variable

whereas steam flow rate, flue gas oxygen concentration, and furnace internal vacuum display more pronounced short-term variability associated with combustion dynamics and control actions. Level-based measurements, including drum and deaerator levels, show quasi-periodic patterns consistent with feedback-regulated operation. Notably, several input variables exhibit clear temporal co-movement with the output power, indicating delayed and coupled system responses rather than instantaneous cause–effect relationships. This multivariate temporal structure confirms the presence of strong sequential dependency and process memory, providing a physically meaningful basis for short-term power output estimation using lag-aware feature engineering and temporally consistent validation strategies.

2.3 Correlation and Linearity Screening

Prior to model training, correlation and linearity analyses were conducted to evaluate inter-variable dependency and potential redundancy within the input space. Such analyses are commonly employed in energy forecasting studies to guide input feature selection and ensure that predictive models rely on variables that are both statistically relevant and physically meaningful [30]. Pearson correlation coefficients were computed between all input variables and the electrical output power. The resulting correlation structure is visualized using a Pearson correlation heatmap, as shown in Fig. 2.

The analysis indicates that steam flow rate exhibits the strongest positive correlation with output power, which is consistent with thermodynamic expectations for Rankine cycle–based biomass power plants. Several boiler-related variables also demonstrate moderate to strong inter-correlations, reflecting the tightly coupled nature of thermo-fluid processes. Variables exhibiting weak linear correlation with output power were retained, as their influence may manifest through nonlinear relationships or delayed system responses rather than direct instantaneous effects.

To further examine the structural form of the input–output relationships, scatter-based linearity screening was performed for the most influential variables, as illustrated in Fig. 3. Steam flow rate shows a clear positive trend with output power; however, the observed dispersion around the regression line suggests that purely linear modeling is insufficient to fully explain output variability. In contrast, variables such as boiler temperature and deaerator level display horizontally clustered distributions with near-zero slopes, confirming their limited direct influence on short-term power modulation. Collectively, these screening results indicate that linear modeling assumptions are inadequate and motivate the adoption of nonlinear learning approaches combined with lag-aware feature design to capture delayed and coupled system dynamics.

2.4 Data Pre-processing, Resampling, and Temporal Dependency Analysis

To ensure numerical stability and comparability across variables with different physical units, all input and output variables were normalized to the [0,1] range using Min–Max scaling. Normalization parameters were derived exclusively from the training data to prevent information leakage during model evaluation [25]. Given the high-frequency nature of the SCADA measurements, multiple resampling strategies were examined to ensure compatibility with different learning algorithms and to balance noise reduction with information preservation. Accordingly, four datasets were constructed:

(i) a high-resolution 1 s dataset using forward-filled imputation to preserve transient dynamics,
(ii) a 1 min mean-resampled dataset to reduce measurement noise while retaining dominant process trends,
(iii) a 5 min mean-resampled dataset primarily used for exploratory visualization, and
(iv) a 1 min median-resampled dataset to improve robustness against sporadic sensor spikes.

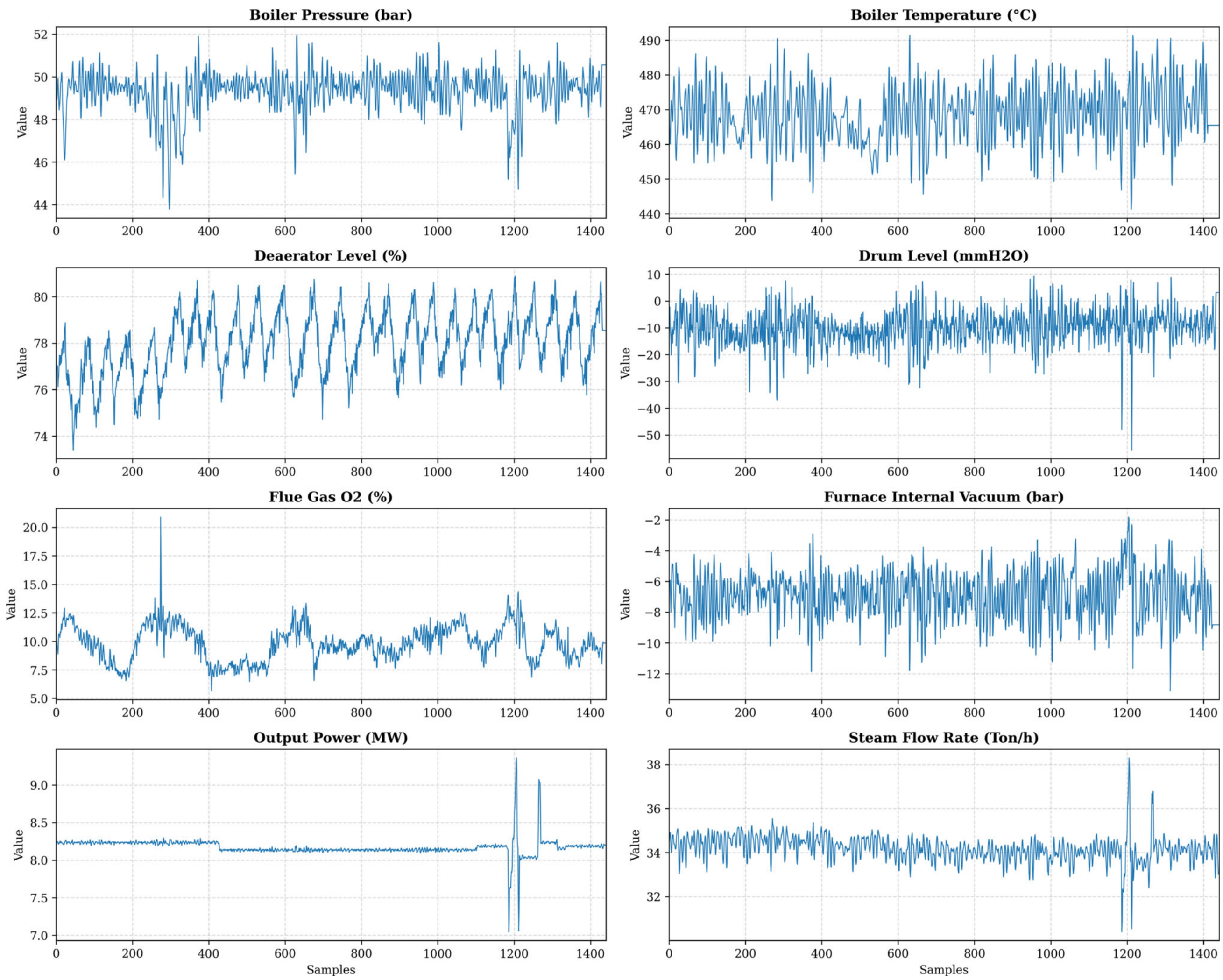

Fig. 1 Multivariate time-series representation of SCADA measurements, including seven process variables and the electrical output power

Across all strategies, the proportion of imputed samples remained below 2%, indicating high data availability and reliable SCADA acquisition.

To assess the impact of temporal aggregation, a temporal reconstruction analysis was performed by comparing the raw 1 s boiler pressure signal with its resampled counterparts at different resolutions. As illustrated in Fig. 4, the 1 min mean and 1 min median resampling strategies preserve the dominant dynamics of the raw signal, whereas 5 min aggregation attenuates short-term transients and suppresses high-frequency variations. Based on this analysis, 1 min mean resampling was selected as the primary pre-processing strategy for subsequent supervised learning, providing a balanced trade-off between noise reduction and temporal fidelity.

Industrial SCADA data inherently exhibit time-series characteristics, where observations at time t are statistically dependent on prior system state $(t-1,\ t-2,\ \ldots)$ [15, 24]. In biomass power plants, such dependencies arise from thermal inertia, transport delays, and closed-loop control mechanisms. To explicitly characterize these temporal dynamics and inform feature design, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analyses were conducted for the output power and selected process variables [28]. The correlograms, evaluated over lag horizons up to 60 min and shown in Fig. 5, indicate strong short-term persistence and delayed system responses.

The observed autocorrelation and partial autocorrelation patterns confirm that the assumption of independent samples is not valid for the considered SCADA time series. Accordingly, lagged representations were incorporated into the model input space. Specifically, lag-1 and lag-2 features were selected to enable ML models to capture delayed and coupled system responses while preserving causal consistency and avoiding the inclusion of future information.

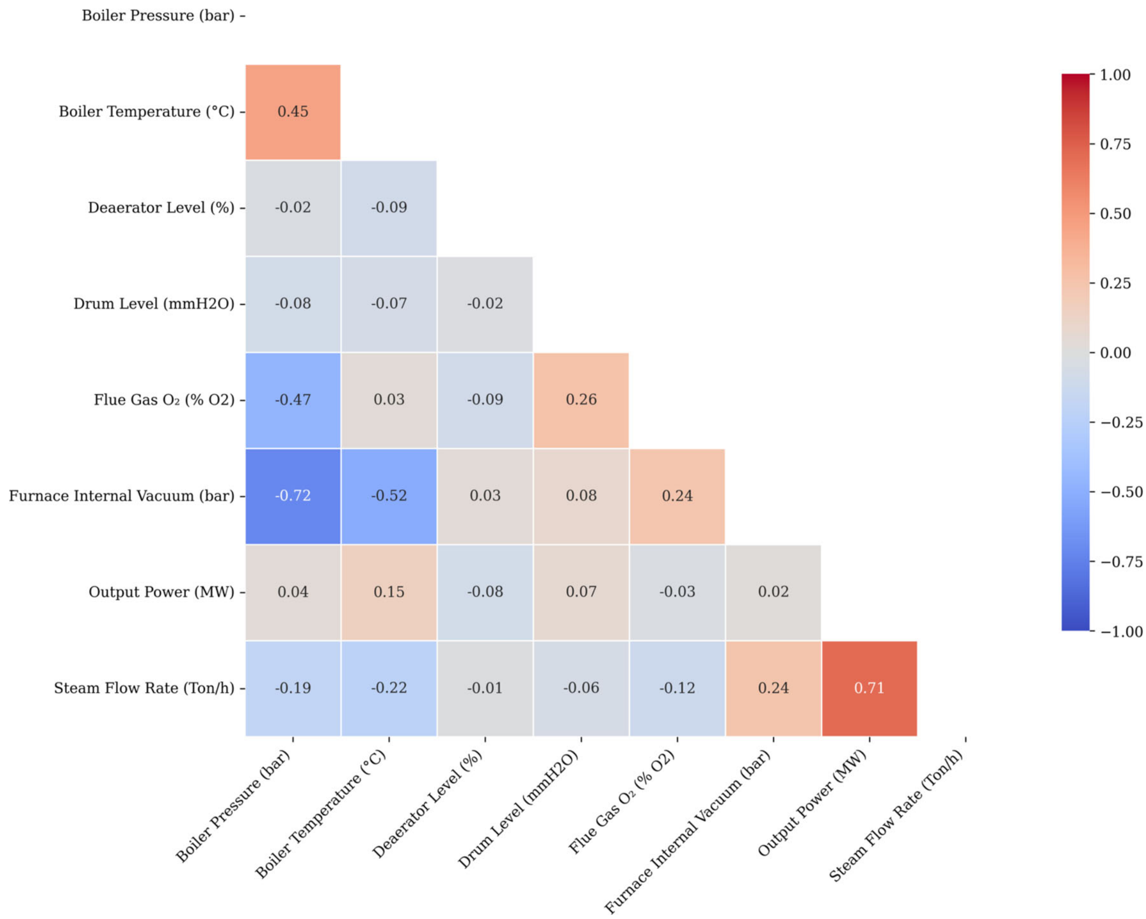

Fig. 2 Pearson correlation heatmap between SCADA variables and electrical output power

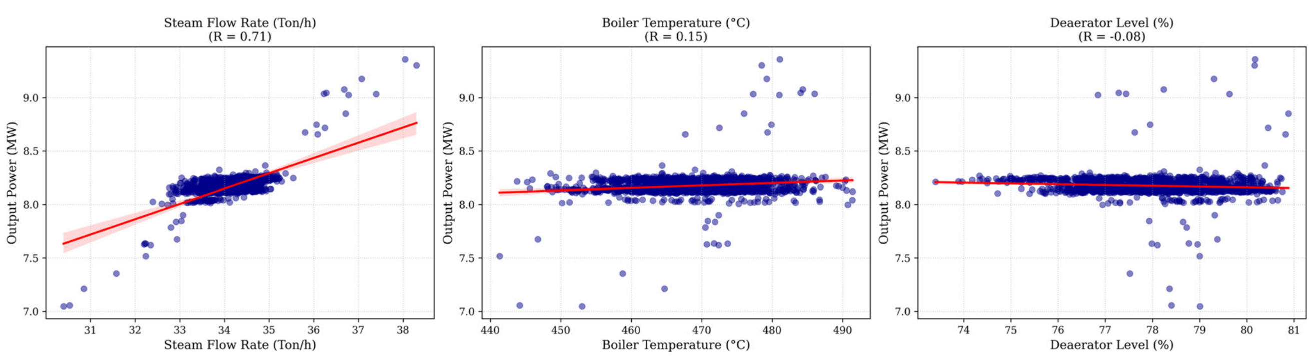

Fig. 3 Scatter plots of selected SCADA variables versus output power, illustrating deviations from linear behavior

2.5 Machine Learning Models

Four supervised ML models were evaluated for regression-based power output estimation: multi-layer perceptron (MLP), random forest (RF), gradient boosting regressor (GBR), and support vector regression (SVR) with a radial basis function kernel. These models were selected due to their demonstrated ability to capture nonlinear input–output relationships in complex energy systems.

MLP represents a widely used baseline neural network model for short-term power prediction tasks [31]. While MLPs are effective in approximating nonlinear mappings, conventional feedforward architectures do not inherently encode temporal dependencies, which can limit their performance in systems characterized by delayed physical responses. RF is an ensemble learning method that combines multiple decision trees trained in parallel through bootstrap sampling and random feature selection, providing

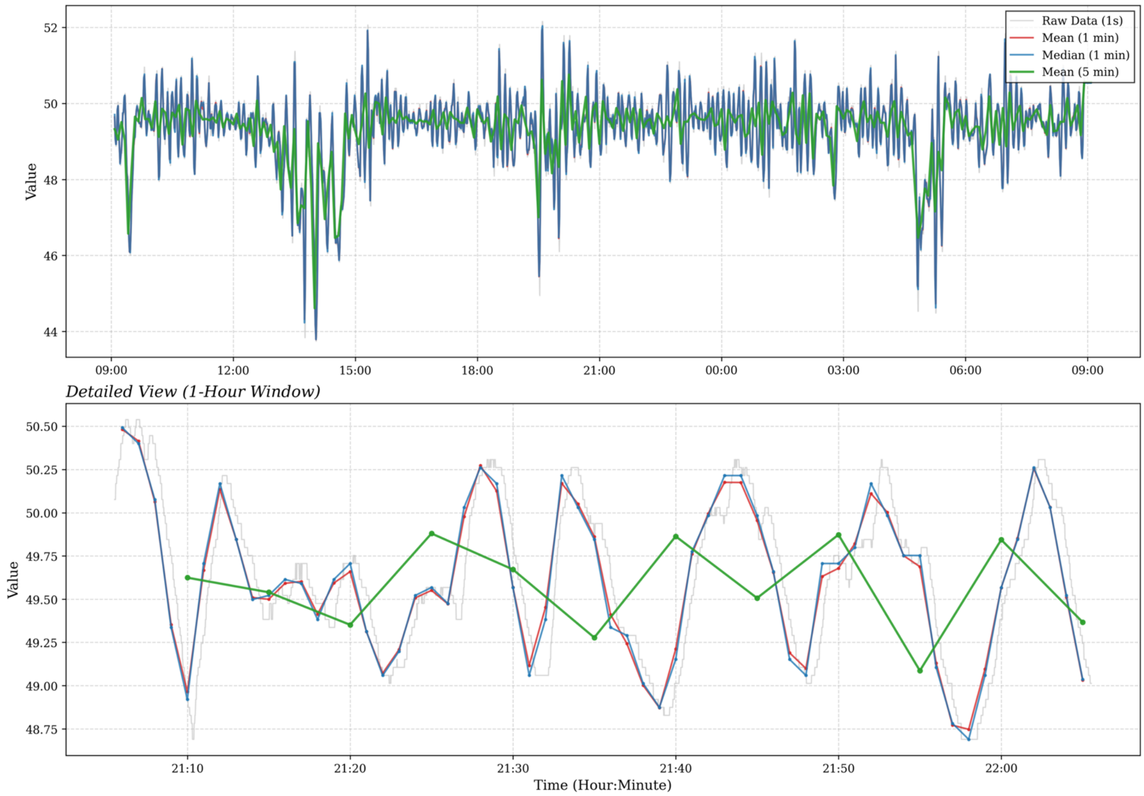

Fig. 4 Time-series reconstruction of Boiler Pressure under different resampling strategies

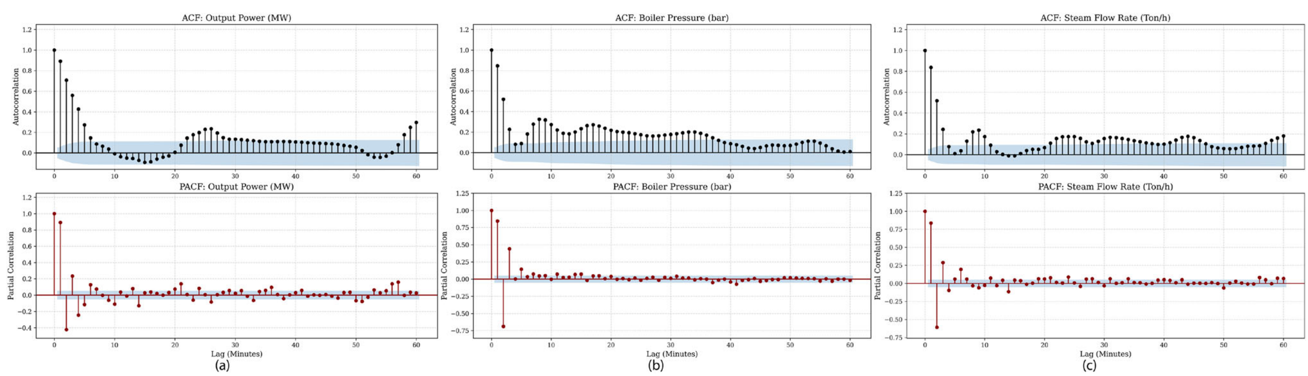

Fig. 5 ACF and PACF analyses of key process variables: **a** Output Power, **b** Boiler Pressure, and **c** Steam Flow Rate

robustness against noise and variance in industrial datasets [21]. However, temporal causality is not explicitly represented in the RF structure, which may reduce predictive reliability in strongly auto-correlated time-series data. In contrast, GBR constructs decision trees sequentially, with each learner correcting the residual errors of its predecessor. This iterative optimization process enables more accurate representation of complex nonlinear and regime-dependent behavior, making GBR particularly suitable for thermodynamically coupled energy conversion processes [21]. SVR employs a kernel-based framework to perform nonlinear regression in a high-dimensional feature space using an ε-insensitive loss function. Although SVR exhibits strong generalization capability, its performance is sensitive to kernel selection and hyperparameter tuning, especially in noisy industrial environments [22]. Together, these models represent complementary neural, ensemble-based, and kernel-based learning paradigms, enabling a systematic comparison of nonlinear regression strategies for biomass power output estimation.

2.5.1 Stage I: Baseline Evaluation

Stage I establishes a non-time-aware baseline using conventional regression-based ML models. In this stage, the SCADA dataset is randomly shuffled and partitioned into training and test subsets using an 80:20 split. This evaluation strategy reflects common practice in data-driven energy modeling, where samples are treated as independent and identically distributed for baseline performance assessment.

All input variables are normalized prior to model training to ensure numerical stability and to avoid dominance effects arising from differing variable scales. The four ML models are trained exclusively on the training subset, while the test subset is reserved for out-of-sample evaluation. Model hyperparameters are optimized using grid search applied solely to the training data to avoid information leakage.

2.5.2 Stage II: Blocked Cross-Validation

In Stage II, the same ML models are evaluated within a time-aware learning framework. While the underlying learning algorithms remain unchanged, temporal consistency is explicitly enforced through data partitioning and feature design. The dataset is divided into training, validation, and test subsets using a blocked splitting strategy composed of contiguous, non-overlapping time segments. This approach preserves chronological order and ensures that models are trained exclusively on historical data and evaluated on future operational periods, thereby maintaining causal consistency [25]. Temporal dependency is incorporated through lag-based feature augmentation. Selected input variables, together with the output power, are expanded with delayed representations at $t-1$ and $t-2$, chosen based on the physical response characteristics of the boiler–steam–turbine system and prior temporal dependency analysis. Hyperparameter optimization is performed using only the training and validation blocks, while the test block remains fully isolated to provide an unbiased assessment of predictive performance under time-consistent conditions.

2.6 Performance Metrics

Reliable evaluation of power output estimation models requires performance metrics that accurately quantify prediction errors while remaining relevant under real industrial operating conditions. Biomass power generation is characterized by continuous-valued outputs and temporal variability arising from thermodynamic inertia, fuel heterogeneity, and closed-loop control actions. Accordingly, model assessment should capture both explanatory capability and absolute prediction error magnitude.

In this study, model performance in both Stage I and Stage II is evaluated using three complementary metrics: the Coefficient of Determination (R^2), Mean Squared Error (MSE), and Mean Absolute Error (MAE), which are widely adopted in short-term power forecasting and renewable energy modeling studies [26]. The R^2 metric measures the proportion of variance in the measured power output explained by the model and provides an indication of goodness of fit.

MSE quantifies the average squared deviation between predicted and measured values, penalizing larger errors more strongly and thereby emphasizing deviations that may have operational consequences. In contrast, MAE represents the average absolute prediction error and offers a more intuitive and directly interpretable measure of typical deviation in power output, expressed in the same units as the target variable.

In Stage I, these metrics are computed on randomly shuffled test samples, corresponding to a conventional evaluation setting commonly reported in data-driven energy modeling studies. In Stage II, the same metrics are calculated on temporally blocked test segments, ensuring causally consistent evaluation in which predictions are generated strictly from past information. This parallel evaluation enables direct assessment of the impact of validation strategy on reported model performance.

In addition to scalar error metrics, time-domain comparisons between measured and predicted power outputs are employed in both stages. These visual analyses provide complementary insight into tracking accuracy and short-term variability that may not be fully captured by aggregate numerical indicators alone. By applying identical performance metrics and visualization criteria across both evaluation stages, the proposed framework enables a fair and systematic comparison between non-time-aware and time-aligned learning configurations.

3 Results and Discussion

3.1 Stage I Results: Baseline Model Comparison

The optimized hyperparameter configurations and corresponding performance metrics obtained for the four ML models (MLP, RF, GBR, and SVR) under Stage I evaluation are summarized in Table 2. For each algorithm, hyperparameters were optimized using grid search to identify the configuration yielding the best test-set performance.

For the RF model, the optimal configuration consists of 100 decision trees with unrestricted tree depth and a minimum split size of five samples. This setting enables the model to capture complex nonlinear interactions in the SCADA data while maintaining a relatively short training time of approximately 11 s. For SVR, the best-performing configuration employs a radial basis function kernel with a regularization parameter C=10 and an ε-insensitive loss width of 0.01.

Table 2 Stage I baseline optimal hyperparameters and test performance of algorithms

Algorithm	Best parameters	Training duration (s)	$\mathbf{R^2}$	MSE	MAE
RF	*n_estimators*: 100, max_depth: None, min_samples_split: 5	10.7	0.0004	0.0134	–
SVR	C: 10, ϵ: 0.01, kernel: rbf	2526.7	0.9001	0.0013	0.0274
GBR	*n_estimators*: 100, max_depth: 5, learning_rate: 0.1	44.0	0.8843	0.0015	0.0298
MLP	hidden_layer_sizes: (7, 30, 30, 30), activation: relu	23.5	0.8492	0.0019	0.0340

Although this configuration improves predictive accuracy relative to simpler kernel settings, it results in substantially higher computational cost, with training times exceeding 2500 s. The GBR achieves its optimal performance using 100 estimators with a maximum tree depth of five and a learning rate of 0.1. Due to its sequential learning mechanism, the corresponding training time is moderately higher (approximately 44 s) than that of RF. For the MLP model, the best results are obtained using a feedforward architecture with four hidden layers structured as (7, 30, 30, 30) and Rectified Linear Unit (ReLU) activation functions, achieving competitive accuracy with a training time of approximately 24 s.

These results highlight a clear trade-off between predictive accuracy and computational efficiency across the evaluated models. Tree-based ensemble methods achieve strong performance with comparatively low training cost, whereas kernel-based learning exhibits significantly higher computational demands when applied to large, high-frequency SCADA datasets. Among the evaluated baseline models, RF attains the highest predictive accuracy with an R^2 value of 0.9687. This performance can be attributed to its ensemble structure and flexible tree depth, which are well suited for modeling nonlinear relationships in tabular process data. SVR achieves the second-highest accuracy (R^2 = 0.9001); however, this performance is accompanied by a substantial computational burden, limiting its scalability in industrial applications [22, 27].

It is important to note that Stage I results are obtained under a random train–test splitting strategy, which is commonly used as a baseline evaluation approach in data-driven energy modeling. In time-dependent datasets, such strategies primarily assess interpolation capability rather than true predictive generalization. As illustrated in Fig. 6, the predicted power signals exhibit noticeable variability and intermittent deviations from the measured trajectory during abrupt operating changes.

This behavior is consistent with prior studies reporting that random sampling in time-ordered data can lead to optimistic performance estimates due to temporal proximity between training and test samples [32–34]. These observations motivate the adoption of time-aware validation strategies, which are examined in Stage II.

3.2 Stage II Results: Blocked Cross-Validation on 1 s Data

When temporal consistency is enforced through blocked cross-validation, the predictive performance of all evaluated models is preserved and, in several cases, improved relative to the baseline evaluation in Stage I. Table 3 summarizes the performance of the ML models under the blocked cross-validation framework using 1 s resolution SCADA data, providing a causally consistent assessment based on time-ordered training and testing.

Among the evaluated approaches, the GBR achieves the highest R^2 (0.9984) using 100 estimators with a maximum tree depth of five and a learning rate of 0.1. This result indicates a very high level of agreement between predicted and measured short-term power variations and reflects the capability of sequential ensemble learning to capture fine-grained nonlinear dynamics in high-frequency biomass power generation data [21, 23]. Owing to its sequential learning strategy, the training time of GBR is approximately 12.8 s, representing a moderate computational cost relative to its achieved accuracy. The RF model attains a comparably high performance (R^2 = 0.9973) with 50 trees, a maximum depth of 20, and a minimum split size of five samples. Due to its parallel tree construction, RF requires only 0.7 s for training, making it the most computationally efficient model under the high-resolution SCADA setting. This favorable balance between predictive accuracy and computational efficiency highlights RF as a strong candidate for real-time or near-real-time deployment, particularly in operational environments with limited computational resources. For the MLP model, the optimal configuration consists of a deep feedforward architecture with hidden layers (7, 30, 30, 30) and ReLU activation functions. When trained on the 1 s dataset, the MLP achieves an R^2 value of 0.9865 with a training time of 0.85 s. This result highlights the sensitivity of neural network performance to temporal resolution and sample density in SCADA-based regression tasks, consistent with prior findings reported in the literature [18]. In contrast, the SVR model, configured with an RBF kernel (C = 10, ε = 0.01), attains a lower accuracy (R^2 = 0.9383) despite benefiting from temporally consistent evaluation. While SVR maintains reasonable generalization capability, its reduced

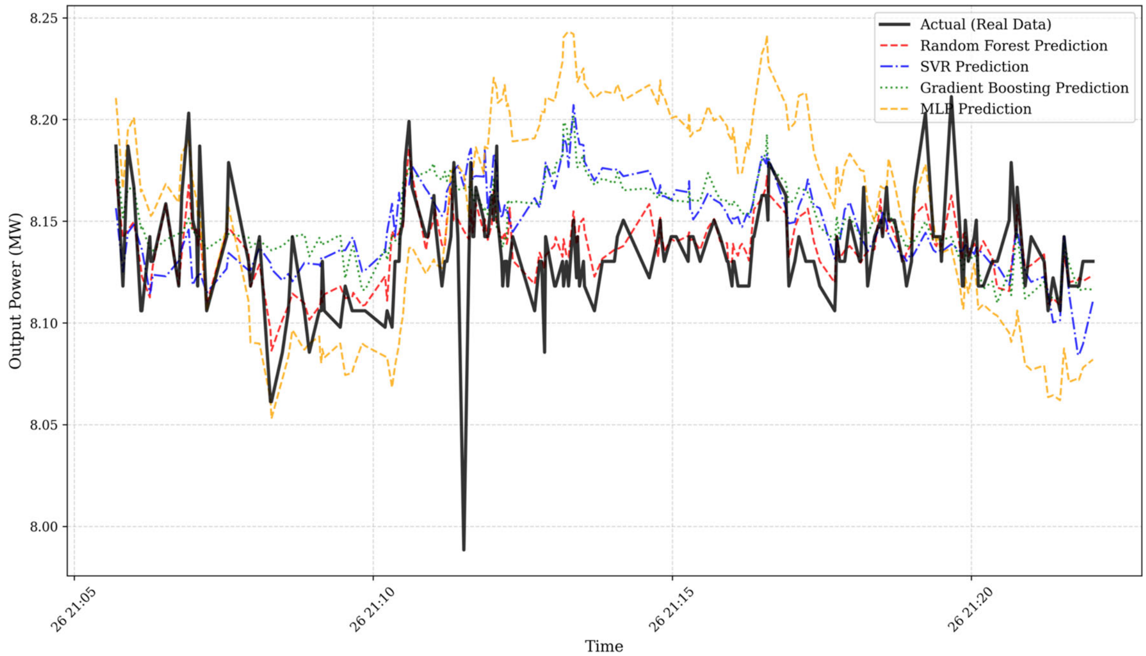

Fig. 6 Time-domain comparison of measured and predicted power output for Stage I models, illustrating prediction variability under random sampling

Table 3 Stage II performance under blocked cross-validation (1 s resolution)

Algorithm	Best parameters	Training duration (s)	R^2	MSE	MAE
GBR	lr: 0.1, depth: 5, n_est: 100	12.80	0.9983	3.75×10^{-6}	0.0010
RF	depth: 20, split: 5, n_est: 50	0.70	0.9973	6.23×10^{-6}	0.0002
MLP	hidden: (7, 30, 30, 30), act: relu	0.85	0.9865	3.12×10^{-5}	0.0036
SVR	C: 10, ϵ: 0.01	0.68	0.9383	1.43×10^{-4}	0.0087

performance suggests limitations in capturing rapid transient dynamics in high-frequency industrial data.

Stage II results demonstrate that combining high-resolution SCADA data with temporally aligned validation enables both improved predictive accuracy and enhanced interpretability of model behavior. The findings further reveal an inherent trade-off between predictive precision and computational efficiency, positioning GBR as the most accurate modeling approach and RF as the most practical solution for operational deployment. While the quantitative metrics reported in Table 3 provide an overall assessment of prediction accuracy, they do not fully reflect how models respond to short-term fluctuations and transient operating conditions. To address this aspect, additional time-domain comparisons are presented to examine temporal consistency and dynamic tracking behavior. Figure 7 illustrates a 200 s zoomed comparison between measured and predicted power outputs for the evaluated models under blocked cross-validation.

3.3 Discussion

A direct comparison between the Stage I and Stage II results demonstrates that evaluation design plays a decisive role in both reported accuracy and temporal robustness of data-driven power estimation models. When random train–test splitting is employed, performance metrics may appear overly optimistic due to the inclusion of temporally adjacent samples across training and test sets. In contrast, blocked cross-validation preserves causal ordering and provides a more conservative yet physically consistent assessment of generalization capability. These findings underline the necessity of time-aware validation for reliable modeling of high-frequency SCADA data in biomass power plants.

The high R^2 observed under blocked validation should be interpreted in the context of the experimental setting. The short forecasting horizon, relatively stable operating regime, and strong auto-correlated structure of the process variables collectively contribute to high predictive performance under

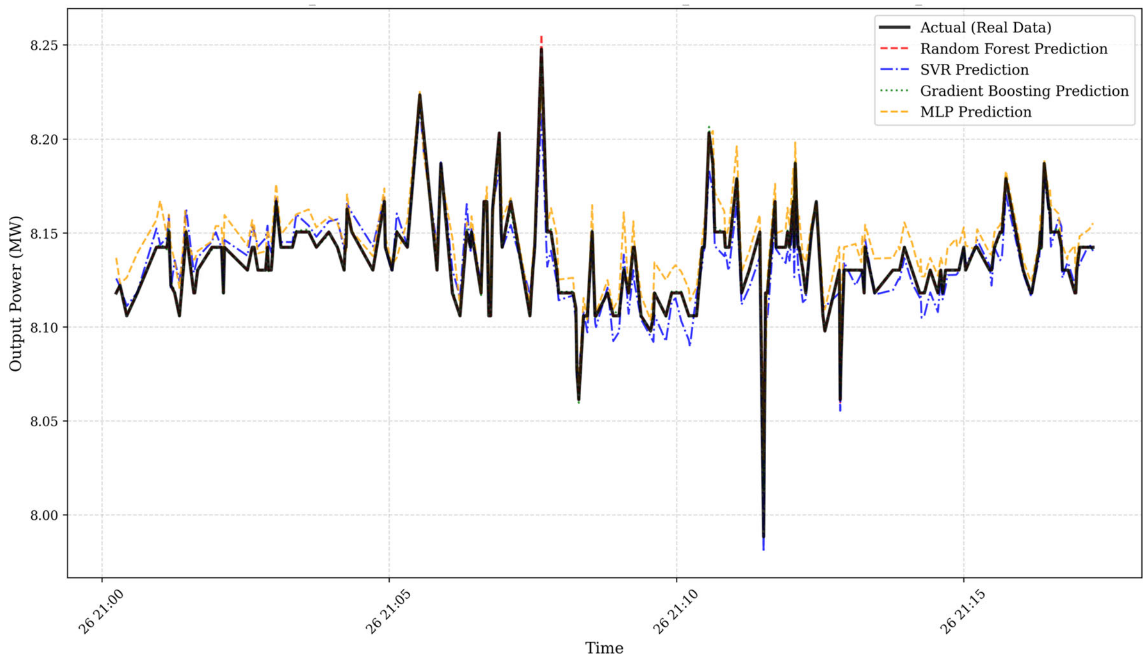

Fig. 7 Comparison of measured and predicted power output over a 200 s test interval for Stage II models, highlighting short-term tracking behavior

controlled conditions. Accordingly, these results should be viewed as indicators of achievable accuracy when temporal consistency is explicitly respected, rather than as universally transferable performance levels.

From a modeling perspective, the results emphasize the fundamental importance of temporal structure in biomass power generation systems. Lag analysis and pre-processing reveal pronounced persistence effects arising from thermal inertia and multi-stage energy conversion processes inherent to biomass combustion and steam-cycle dynamics. These characteristics render instantaneous regression assumptions physically insufficient for short-term power estimation using high-resolution SCADA data.

Across both evaluation stages, ensemble-based models, particularly GBR and RF, consistently exhibit strong predictive performance, reflecting their ability to capture nonlinear interactions and regime-dependent responses in tabular process data. In contrast, the MLP model shows marked sensitivity to the validation strategy. Its substantial performance improvement under blocked evaluation indicates that neural network models are particularly susceptible to misleading performance estimates when inappropriate random sampling schemes are applied.

The performance gains observed under temporally blocked validation do not indicate overfitting. Rather, they result from the elimination of temporal information leakage inherent to random sampling strategies that violate causal ordering. Enforcing temporal continuity enables a more physically meaningful mapping between process variables and electrical power output, leading to improved stability, interpretability, and reliability of model predictions.

The findings demonstrate that temporally aligned learning frameworks enhance both accuracy and robustness in short-term biomass power output estimation. When appropriately configured, data-driven models exhibit strong potential for deployment in real-time monitoring, anomaly detection, and decision-support applications in industrial biomass power plants. Nevertheless, the present study is limited to a single facility operating under relatively stable conditions.

4 Conclusions

This study addressed the problem of short-term power output estimation in an industrial biomass power plant by explicitly accounting for the temporal structure of high-frequency SCADA data. A two-stage ML framework was proposed to systematically compare conventional random sampling–based evaluation with a temporally consistent blocked validation strategy across four nonlinear regression models. The results demonstrate that random train–test splitting leads to overly optimistic performance estimates due to temporal information leakage, whereas time-aware validation yields more physically consistent and reliable predictions. By integrating lag-aware feature engineering with causally aligned evaluation, the proposed approach provides a robust and transparent basis for assessing ML model performance in biomass power generation systems.

From a model-specific perspective, the results indicate that ensemble-based ML models offer the most reliable performance for short-term power output estimation in biomass power plants when temporal consistency is enforced. GBR achieves the highest overall accuracy, demonstrating superior capability in capturing nonlinear interactions and delayed process dynamics, which is particularly advantageous for high-resolution monitoring applications requiring maximum precision. RF provides a closely comparable level of accuracy while offering significantly lower computational cost, making it a more suitable candidate for real-time deployment scenarios and adaptive retraining environments. In contrast, the MLP model exhibits strong sensitivity to data volume and evaluation design; its substantial performance improvement under temporally blocked validation confirms that neural network–based models benefit markedly from high-frequency data but are prone to misleading assessments under inappropriate sampling strategies. SVR, although effective under certain conditions, shows limited scalability for large SCADA datasets due to its computational complexity. Collectively, these findings highlight that model selection in industrial energy systems should be guided not only by aggregate accuracy metrics but also by temporal robustness, computational efficiency, and deployment constraints.

Future work may extend the proposed time-aware evaluation framework to biomass power plants operating under different fuel compositions, seasonal conditions, and load regimes to assess robustness and generalizability. Incorporating longer forecasting horizons and adaptive retraining strategies could enable evaluation under more dynamically varying operational scenarios. In addition, the integration of sequence-based DL architectures and hybrid physics-informed learning approaches represents a promising direction for capturing longer-term dependencies while preserving physical interpretability. These extensions are expected to further enhance the applicability of temporally consistent ML models for real-time monitoring, anomaly detection, and decision-support systems in industrial biomass power generation.

Acknowledgements The authors thank SATEM SİNOP BİYOKÜTLE ENERJİ Ltd. ŞTİ. management for providing operational data used in this study. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

References

1. Demirbas, A.: Importance of biomass energy sources for Turkey. Energy Policy **36**(2), 834–842 (2008). https://doi.org/10.1016/j.enpol.2007.11.005
2. Dincer, I.; Acar, C.: Smart energy systems for a sustainable future. Appl. Energy **194**, 225–235 (2017). https://doi.org/10.1016/j.apenergy.2016.12.058
3. Agency, I.E.: Renewables 2024: analysis and forecast to (2030). https://www.iea.org/reports/renewables-2024
4. Mignogna, D.; Szabó, M.; Ceci, P.; Avino, P.: Biomass energy and biofuels: perspective, potentials, and challenges in the energy transition. Sustainability **16**(16), 7036 (2024). https://doi.org/10.1002/bbb.70106
5. Werther, J.; Saenger, M.; Hartge, E.-U.; Ogada, T.; Siagi, Z.: Combustion of agricultural residues. Progress Energy Combust. Sci. **26**(1), 1–27 (2000). https://doi.org/10.1016/S0360-1285(99)00005-2
6. Nussbaumer, T.: Combustion and co-combustion of biomass: fundamentals, technologies, and primary measures for emission reduction. Energy Fuels **17**(6), 1510–1521 (2003). https://doi.org/10.1021/ef030031q
7. Baxter, L.: Biomass-coal co-combustion: opportunity for affordable renewable energy. Fuel **84**(10), 1295–1302 (2005). https://doi.org/10.1016/j.fuel.2004.09.023
8. Obernberger, I.; Thek, G.: Physical characterisation and chemical composition of densified biomass fuels with regard to their combustion behaviour. Biomass Bioenergy **27**(6), 653–669 (2004). https://doi.org/10.1016/j.biombioe.2003.07.006
9. Jones, J.; Darvell, L.; Bridgeman, T.; Pourkashanian, M.; Williams, A.: An investigation of the thermal and catalytic behaviour of potassium in biomass combustion. Proce Combust Inst **31**(2), 1955–1963 (2007). https://doi.org/10.1016/j.proci.2006.07.093
10. Taye, M.M.: Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. Computers **12**(5), 91 (2023). https://doi.org/10.3390/computers12050091
11. Catalão, J.D.S.; Pousinho, H.M.I.; Mendes, V.M.F.: Short-term wind power forecasting in Portugal by neural networks and wavelet transform. Renew Energy **36**(4), 1245–1251 (2011). https://doi.org/10.1016/j.renene.2010.09.016
12. Yamin, H.; Shahidehpour, S.; Li, Z.: Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets. Int J Electr Power Energy Syst **26**(8), 571–581 (2004). https://doi.org/10.1016/j.ijepes.2004.04.005
13. Sopeña, J.M.G.; Pakrashi, V.; Ghosh, B.: A benchmarking framework for performance evaluation of statistical wind power forecasting models. Sustain Energy Technol Assess **57**, 103246 (2023). https://doi.org/10.1016/j.seta.2023.103246
14. Kusiak, A.; Zheng, H.; Song, Z.: Wind farm power prediction: a data-mining approach. Wind Energy Int J Progress Appl Wind Power Conversion Technol **12**(3), 275–293 (2009). https://doi.org/10.1002/we.295
15. Mafakheri, F.; Nasiri, F.: Modeling of biomass-to-energy supply chain operations: applications, challenges and research direc-

tions. Energy Policy **67**, 116–126 (2014). https://doi.org/10.1016/j.enpol.2013.11.071

16. 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 Measur Control **43**(1), 78–87 (2021). https://doi.org/10.1177/0142331220932390
17. Yang, Q.; Zhang, J.; Zhou, J.; Zhao, L.; Zhang, D.: A hybrid data-driven machine learning framework for predicting the performance of coal and biomass gasification processes. Fuel **346**, 128338 (2023). https://doi.org/10.1016/j.fuel.2023.128338
18. Benti, N.E.; Chaka, M.D.; Semie, A.G.: Forecasting renewable energy generation with machine learning and deep learning: current advances and future prospects. Sustainability **15**(9), 7087 (2023). https://doi.org/10.3390/su15097087
19. Toth, P.; Garami, A.; Csordas, B.: Image-based deep neural network prediction of the heat output of a step-grate biomass boiler. Appl Energy **200**, 155–169 (2017). https://doi.org/10.1016/j.apenergy.2017.05.080
20. Kotyra, A.; Gromaszek, K.: Determination of coal and biomass co-combustion process states using convolutional neural networks. Energies **18**(19), 5219 (2025). https://doi.org/10.3390/en18195219
21. Brandić, I.; Pezo, L.; Bilandžija, N.; Peter, A.; Šurić, J.; Voća, N.: Comparison of different machine learning models for modelling the higher heating value of biomass. Mathematics **11**(9), 2098 (2023). https://doi.org/10.3390/math11092098
22. Bilgiç, G.; Gök, A.E.: Comparison of random forest and support vector regression models in predicting hydrogen production process from biomass. Çukurova Üniversitesi Mühendislik Fakültesi Dergisi **39**(2), 475–488 (2024). https://doi.org/10.21605/cukurovaumfd.1514518
23. Elmaz, F.; Yücel, Ö.; Mutlu, A.Y.: Predictive modeling of biomass gasification with machine learning-based regression methods. Energy **191**, 116541 (2020). https://doi.org/10.1016/j.energy.2019.116541
24. Pandit, R.; Astolfi, D.; Hong, J.; Infield, D.; Santos, M.: Scada data for wind turbine data-driven condition/performance monitoring: a review on state-of-art, challenges and future trends. Wind Eng **47**(2), 422–441 (2023). https://doi.org/10.1177/0309524X221124031
25. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W., et al.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography **40**(8), 913–929 (2017). https://doi.org/10.1111/ecog.02881
26. Hyndman, R.J.; Athanasopoulos, G.: Forecasting: principles and practice. OTexts, Melbourne (2018)
27. Yang, C.-T.; Kristiani, E.; Leong, Y.K.; Chang, J.-S.: Big data and machine learning driven bioprocessing-recent trends and critical analysis. Bioresource Technol **372**, 128625 (2023). https://doi.org/10.1016/j.biortech.2023.128625
28. Cerqueira, V.; Torgo, L.; Mozetič, I.: Evaluating time series forecasting models: an empirical study on performance estimation methods. Mach Learn **109**(11), 1997–2028 (2020). https://doi.org/10.1007/s10994-020-05910-7
29. Kaufman, S.; Rosset, S.; Perlich, C.; Stitelman, O.: Leakage in data mining: formulation, detection, and avoidance. ACM Trans Knowl Discov Data (TKDD) **6**(4), 1–21 (2012). https://doi.org/10.1145/2382577.2382579
30. Jebli, I.; Belouadha, F.-Z.; Kabbaj, M.I.; Tilioua, A.: Prediction of solar energy guided by pearson correlation using machine learning. Energy **224**, 120109 (2021). https://doi.org/10.1016/j.energy.2021.120109
31. Haykin, S.: Neural Netw Learn Mach, 3rd edn. Pearson Education India, New Delhi (2009)
32. Bergmeir, C.; Benítez, J.M.: On the use of cross-validation for time series predictor evaluation. Inform Sci **191**, 192–213 (2012). https://doi.org/10.1016/j.ins.2011.12.028
33. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; Warton, D.I.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography **40**(8), 913–929 (2017). https://doi.org/10.1111/ecog.02881
34. Cerqueira, V.; Torgo, L.; Pinto, F.; Soares, C.: Evaluating time series forecasting models: an empirical study on performance estimation methods. Mach Learn **109**(11), 1997–2028 (2020). https://doi.org/10.1007/s10994-020-05910-7