Energy Conversion and Management 357 (2026) 121464

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Research Paper

Bi-directional long short-term memory based ensemble deep learning framework for non-linear steam turbine power forecasting: a biomass fuelled case study

Himaya Perera [a,1], Shalitha Jayasekara [b,1], Ruchire Eranga Wijesinghe [c,e], Bhagya Nathali Silva [d,e,*], Honnyong Cha [b]

[a] School of Computing, Engineering and Mathematical Sciences, La Trobe University, Bundoora, Victoria, Australia
[b] School of Energy Engineering, Kyungpook National University, Daegu, Republic of Korea
[c] Department of Electrical and Electronic Engineering, Faculty of Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka
[d] Department of Information Technology, Faculty of Computing, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka
[e] Centre for Excellence in Informatics, Electronics & Transmission (CIET), Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

ARTICLE INFO

Keywords:
BiLSTM
Steam turbine optimization
Biomass energy optimization
Power forecasting
Time-series forecasting

ABSTRACT

In palm oil manufacturing, steam turbines powered by biomass fuel are central to energy generation. However, fluctuating load demands and temporal variations lead to inefficiencies, while limited and variable supply of biomass waste constrains boiler feed flexibility. Current index-based boiler feeding methods overlook actual load demands and waste availability, resulting in significant energy wastage. This study presents a novel ensemble deep learning model combining Bidirectional Long Short-Term Memory (Bi-LSTM) and Gated Recurrent Units (GRU) with Attention Layers, trained on an eight-year operational dataset with structured preprocessing and feature selection, to forecast steam turbine power generation. The model captures complex non-linear temporal patterns more effectively than conventional and standalone ML models, achieving a Root Mean Square Error (RMSE) of 0.0684, Mean Absolute Error (MAE) of 0.0414, and an R-squared (R^2) value of 0.9832, which outperformed eight benchmark models by approximately 25% in prediction accuracy. Additionally, the framework incorporates operational parameters such as kVA, total energy, and Fresh Fruit Bunch (FFB) production to dynamically optimise biomass feed rates, balancing energy output with resource availability. This approach minimises energy wastage, reduces grid reliance, and promotes both sustainability and profitability.

1. Introduction

Palm oil is currently the world's most widely produced edible oil. The palm oil manufacturing industry is integral to global agriculture, food production, pharmaceuticals, cosmetics, and personal care production, contributing significantly to economic development [1] and renewable energy initiatives by utilising its production waste. In this context, biomass waste from palm fruit fibres, shells, and husks are used as a fuel source for boilers to generate steam power [2] in palm oil manufacturing companies. However, the fluctuating load demands driven by production schedules, order volumes, and time-based variations pose challenges to efficient power generation [3]. The inconsistency in energy demand often leads to periods of excess energy production resulting in significant energy wastage. Additionally, the mass flow to the boiler is inherently constrained by the availability of biomass waste from palm oil production processes, further limiting the flexibility of energy generation. Currently, the boiler is fed according to a fixed hourly index without considering actual energy demands or wastage in palm oil manufacturing. This approach is inefficient, especially given the limited availability of biomass fuel. During peak demand periods, this fixed system can lead to fuel shortages and reliance on costly grid energy, while during low-demand periods, it can result in overproduction and wasted energy. These inefficiencies underscore the need for a more adaptive power forecasting model to account for load variability, optimise energy utilisation, and improve production waste management. The manufacturing process can be more sustainable and profitable by dynamically adjusting fuel feed rates and increasing reliance on steam

* Corresponding author at: Department of Information Technology, Faculty of Computing, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka.
E-mail address: nathali.s@sliit.lk (B.N. Silva).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.enconman.2026.121464
Received 10 May 2025; Received in revised form 16 March 2026; Accepted 6 April 2026
Available online 10 April 2026

Nomenclature

Abbreviation	Definition
AFR	Air-Fuel Ratio
ANNs	Artificial Neural Networks
ARMA	Autoregressive Moving Average
Bi-GRU	Bidirectional Gated Recurrent Unit
CHP	Combined Heat and Power
CNNs	Convolutional Neural Networks
D-GRU	Drop Gated Recurrent Unit
DBN	Deep Belief Network
ECLIPSE	European Coal Liquefaction Process Simulation and Evaluation
FFBs	Fresh Fruit Bunches
FiT	Feed-in Tariffs
GA	Genetic Algorithm
GRU	Gated Recurrent Unit
ICA	Imperialist Competitive Algorithm
KS	Kernel Shells
LP steam	Low-Pressure Steam
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MFE	Mean Forecast Error
MFES	Multi-Factor Evolution Strategy
MPC	Model Predictive Control
NLP	Non-linear Programming
NN	Neural Network
PID	Proportional-Integral-Derivative
POME	Palm Oil Mill Effluent
PSO	Particle Swarm Optimisation
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RSM	Response Surface Methodology
SA	Simulated Annealing
SD	Standard Deviation
SVMs	Support Vector Machines
TFT	Temporal Fusion Transformer
XGBoost	eXtreme Gradient Boosting

turbine-generated power.

To address these challenges, current approaches in palm oil manufacturing focus on a combination of cogeneration systems [4], enhanced biomass utilisation strategies, and supportive policy frameworks. Cogeneration, which simultaneously produces electricity and thermal energy, is widely adopted in palm oil mills. Biomass residues like palm kernel shells, fibres, and empty fruit bunches (EFBs) [5] are used in boilers to generate steam for both electricity production and processing. However, EFBs, with high moisture content often require pre-treatment as shown in Fig. 1 for efficient combustion, which can limit their use as a consistent fuel source.

In regions like Malaysia and Indonesia, governmental policies, such as the FiT (Feed-in tariffs) system, incentivize mills to feed excess electricity generated from biomass into the grid, thus reducing energy waste and promoting renewable energy use. Another emerging solution is the anaerobic digestion of palm oil mill effluent (POME) to produce biogas, which can power internal combustion engines or micro-turbines, supplementing electricity generation from solid biomass [6]. Research also emphasizes the potential of clustering biomass power plants, particularly in areas with dense palm oil mill operations. By minimising transport costs and integrating biomass resources across mills, this approach allows for more efficient energy management and reduces reliance on fossil fuels in rural areas. These solutions collectively aim to stabilize power output, match generation with demand, and minimise the environmental impact of palm oil production. Palm oil mills primarily use combined heat and power (CHP) [7] systems, efficiently generating electricity and heat from biomass residues like palm kernel shells and mesocarp fibres. This approach optimises available energy by providing power for the mill’s operations and utilising excess heat for processes like sterilizing fresh fruit bunches (FFBs). Some approaches focus on modelling and optimising biomass-based cogeneration plants, with key simulations conducted using software such as Aspen Plus [8]. Key techniques include evaluating airflow rates and biomass flow to determine the optimal air-fuel ratio (AFR), which improves combustion conditions and stabilizes power output. Various scenarios, including low-pressure steam (LP-steam) recycling and the use of exhaust heat, have been simulated to boost the system's efficiency. Implementing LP-steam and condensate recycling increases the recycling capacity up to 80% with exhaust utilization and enhances the energy yield by maximizing heat recovery from exhaust gases. Wu et al.[9] utilised ECLIPSE software to optimise Combined Heat and Power (CHP) systems by analysing fuel combinations such as EFBs, kernel shells (KS), and biogas for enhanced power generation efficiency. Preheating techniques for inlet air and water were also employed, to significantly improve combustion efficiency and power output.

In recent years, the focus has progressively shifted from conventional thermodynamic models [10] to advanced [11,12], AI-driven methods to enhance the efficiency of steam turbine power generation. Traditional approaches, such as the Willans line model [10,13], have laid the groundwork for turbine performance analysis but are often constrained

Fig. 1. Pre-treatment process of empty fruit bunches. a) harvested fruit bunches, (b) de-fibred bunches, and (c) processed fibres.

by high computational demands and reliance on multiple fixed constants and parameters. Thermodynamic optimisation techniques like intercoolers, regenerators, and economizers have been introduced to improve the thermodynamic performance of the Brayton Cycle, complementing those traditional models [14]. These systems enhance energy recovery and conversion efficiency, addressing the limitations of conventional methods. Furthermore, advancements in control strategies, such as Proportional-Integral-Derivative (PID) [15] controllers, have offered simple yet effective solutions to correct steady-state errors in turbine operations. More sophisticated approaches like Model Predictive Control (MPC) [15,16] have emerged, enabling multi-variable control system optimisation while managing operational constraints. This dynamic adaptability of MPC ensures turbine efficiency under fluctuating load conditions. Together, these innovations signify a paradigm shift, blending traditional thermodynamic principles with cutting-edge computational methods to meet the demands of modern steam turbine power generation.

Hochreiter [17] demonstrated that Long Short-Term Memory (LSTM) networks capture time-dependent and complex temporal patterns, making them essential for precise power forecasting. Furthermore, hybrid methods [18] that combine LSTM with other advanced neural network architectures, like Convolutional Neural Networks (CNNs) [19], yield even greater accuracy and resilience in predicting power generation dynamics. Some approaches include the use of

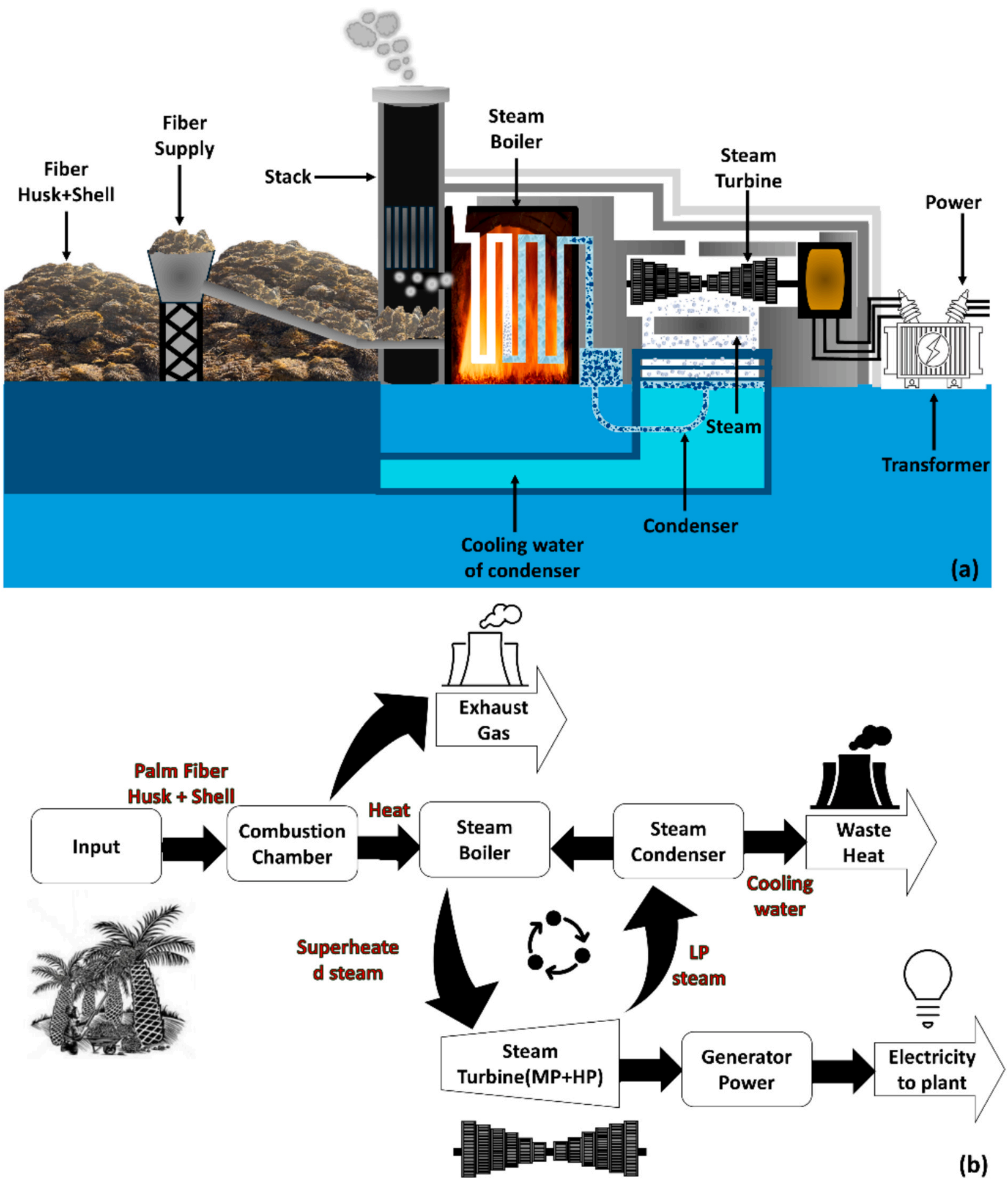

Fig. 2. Schematic of steam turbine operation utilising palm biomass integrated with the combustion, boiler, and turbine system for power generation.

Artificial Neural Networks (ANNs) [20] and Support Vector Machines (SVMs) [21] to model and optimise turbine performance under varying operational conditions. These models allowed precise adjustments of operating parameters, such as steam temperature and pressure, to maximize isentropic efficiency and minimize emissions. Additionally, Monte Carlo-based [22] sensitivity analysis is used to identify the most impactful parameters, further refining optimisation efforts. Hybrid frameworks, combining AI models with traditional optimisation techniques like Non-linear Programming (NLP), facilitate real-time adjustments that improve efficiency under different load scenarios. In the domain of turbine power prediction, prior studies have extensively explored both conventional and AI-based methodologies [4,10,23]. Conventional approaches, such as thermodynamic models and time-series statistical techniques, offer simplicity and lower computational overhead but fall short in capturing complex non-linear dynamics and adapting to fluctuating load conditions. AI-based methods, including ANNs, LSTM networks, and SVMs, have significantly improved in predictive accuracy by leveraging advanced data-driven techniques. Hybrid models [24,6], which combine complementary methods like LSTM and CNNs, further enhance performance by integrating temporal and spatial feature extraction. However, these hybrid approaches still struggle with real-time adaptability and robustness under highly dynamic operational environments, limiting their scalability and generalizability across varied conditions.

This research introduces an innovative ensemble methodology in recognizing those challenges [7] that integrates multiple advanced AI architectures. Unlike traditional single or hybrid models, this ensemble approach combines the strengths of various architectures, effectively addressing the gaps in accuracy, resilience, and adaptability highlighted in prior studies. The comparison of existing models, as outlined in the accompanying table, underscores the unique advantages of this ensemble framework. By optimising the predictive accuracy and robustness of power forecasting, the proposed model minimises errors and supports adaptive, real-time decision-making in turbine operations, while advancing sustainability by addressing inefficiencies and leveraging predictive models.

A biomass-fuelled steam turbine scenario illustrated in Fig. 2(a) has been used as the case study for this research. The process depicted in the Fig. 2(b) highlights the flow of palm oil biomass residues, such as husks and shells, which serve as fuel in the combustion chamber. This systematic representation of the biomass-to-power cycle forms the foundation for analysing and optimising power generation in palm oil mills.

The below Table 1 includes a summary of various models and techniques employed for power prediction and forecasting, highlighting their methodologies, applications, and key performance metrics across different systems.

Table 1
Summary of Models and Techniques for Steam Turbine Performance.

References	Model	Type	Objectives	Pros	Research Gap
[10]	Thermodynamic Models (Willans Line)	Conventional	Predict steam turbine performance	Simple, low computational cost	Challenges in Data Regression and Optimisation, lack of real-time adaptability
[22]	Monte Carlo Sensitivity Analysis	Statistical	Identify key parameters for optimisation	Useful for risk assessment	Computationally intensive, lacks predictive capabilities
[25]	ARMA, ARIMA Models	Statistical	Short-term power prediction for linear relationships	Good for small-scale data; interpretable	Limited to linear relationships, unable to handle complex, non-linear dependencies
[26]	Artificial Cooperative Search (ACS)	Metaheuristic Optimisation	Forecast electric energy consumption using socio-economic indicators	High accuracy and reliability in forecasting, Superior to GA, PSO, SA, ICA, etc.	Scalability to other regions or applications does not explore
[26]	Linear, Quadratic, Exponential, Logarithmic Models	Statistical Models	Optimise weighting factors for energy consumption models	Foundational basis for optimisation algorithms	Limited flexibility for non-linear relationship
[27]	Deep Belief Network (DBN)	AI-based	Forecast short-term power load (day-ahead and week-ahead)	High accuracy in short-term forecasting, mitigates overfitting.	High dependence on hyperparameters, limited scalability for long-term forecasts
[28]	SARIMA-GARCH with Skew-Normal Distribution	Statistical Time Series	Predicts short-term electricity load for 24 h	Lower square error and absolute error compared to ARIMA-GARCH, handles seasonal impacts, global trends, and heteroscedasticity.	Focuses on short-term (24-hour) forecasting only, lacking scalability for longer-term predictions.
[28]	ARIMA-GARCH with Normal Distribution	Statistical Time Series	Baseline comparison for load prediction	Handles conditional heteroscedasticity and capturing linear/non-linear correlations.	Lower accuracy compared to SARIMA-GARCH with Skew-Normal Distribution
[29]	Wavelet CS-HANFIS, Wavelet GSA-HANFIS, Wavelet COA-HANFIS	Neuro-Fuzzy with Heuristics	Monthly electricity demand forecasting	Best fit for historical data, Combines wavelet transform, neuro-fuzzy systems, and heuristic optimisation algorithms.	Limited application scope, focused on one heuristic algorithm, requires comparative performance validation in different scenarios.
[30]	Combined Method (BP + ANFIS + Diff-SARIMA)	Hybrid Statistical & AI	Short-term electricity demand forecasting	Combines linearity, non-linearity, and seasonality data effectively.	Limited validation against other advanced hybrid methods beyond MFES.
[20]	ANN RSM	AI-based Statistical	Predict turbine performance under varying conditions	Handles non-linearities well	Limited to short-term trends, lacks robustness to large fluctuations, Modelling Limitations
[10]	LSTM	AI-based	Time-series forecasting	Captures long-term dependencies	Limited Applicability of Conventional LSTM
[21]	SVMs	AI-based	Predict turbine efficiency based on input parameters	High accuracy for small datasets, Effective for Stochastic System	Poor scalability to larger datasets, struggles with multi-dimensional data
[31]	CNN-LSTM	AI-based	Predict short-term photovoltaic (PV) power production over different look-back and look-forward windows.	Combines CNN's feature extraction with LSTM's temporal learning.	Focus on Short-Term Forecasting
[6]	Hybrid LSTM-CNN	AI-based	Enhance power forecasting accuracy	Combines temporal and spatial feature extraction	Complexity in Data Processing.

Based on the identified research gaps and practical needs in biomass-based power forecasting, the key objectives of this study are defined as follows: (See Table 2).

- To develop a forecasting model that captures the non-linear, volatile dynamics of steam turbine power generation in palm oil manufacturing
- To systematically benchmark the proposed model against eight conventional, hybrid, and ensemble ML baselines.
- To optimize forecasting accuracy using structured hyperparameter tuning and chronological validation.
- To evaluate real-world applicability using an industrial dataset.
- The uniqueness of our work lies in the integration of Bi-LSTM, Bi-GRU, and Attention layers into a unified ensemble framework, which achieves superior accuracy (25% improvement over baselines) while maintaining computational efficiency for industrial deployment.

Section 2 details the proposed methodology. Section 3 presents analyses the results. Section 4 discusses the findings and key implications. Finally, Section 5 concludes the study, highlighting contributions, limitations, and future research directions.

2. Methodology

2.1. Workflow overview

The workflow for this research is outlined in Fig. 3(a). It begins with data collection. Then after data pre-processing, missing values were addressed using statistical methods such as mean imputation, and hourly data was aggregated into daily observations to reduce noise and capture long-term trends. Feature selection was performed to identify the most relevant parameters impacting power generation. Following this, the neural network architecture was designed to capture complex temporal patterns in the data. The model was then trained using the selected features, followed by a rigorous evaluation process based on standard performance metrics to ensure reliability and precision. Finally, the most effective model was selected to generate future power predictions, supporting improved operational efficiency and decision-making in steam turbine systems.

2.2. Data Collection and Input Features

The dataset was collected from a steam turbine operational logbook of a palm oil manufacturing company, spanning for eight years (2016–2023) and included approximately 70,000 hourly observations recorded in the plant logbook. For modelling purposes, these were aggregated into ~2,920 daily records to emphasize long-term forecasting patterns while reducing short-term noise. The data comprises critical parameters recorded during the operation of biomass-fuelled steam turbines, which are integral to the plant's power generation process. Palm fruit residues (fibers, shells, and empty fruit bunches) served as the consistent biomass feedstock throughout the study period. These residues are widely reported in literature to have an average calorific value of 15–18 MJ/kg, which corresponds to the typical energy content utilized in palm oil mill steam turbines. Initially, a wide range of features were logged, including turbine speed, inlet steam pressure, nozzle steam pressure, exhaust steam pressure, lubrication oil pressures and temperatures, bearing temperatures, voltage, current, and generated power. An initial analysis was conducted to examine the correlation coefficients between input features and the power generation to develop a robust and accurate forecasting model. Based on the correlation analysis illustrated in Fig. 4, Turbine Speed, Inlet Steam Pressure, Nozzle Steam Pressure, Exhaust Steam Pressure, and KVA Value were identified as critical factors affecting power generation. Consequently, they were selected for further analysis and training, ensuring that the forecasting model leveraged the most impactful predictors while minimising redundancy and improving computational efficiency.

Table 2
Performance Comparison of Different Models.

Model	RMSE ± SD	R^2 ± SD	MAE ± SD	MAPE (%) ± SD	Bias (MFE) ± SD
LSTM	0.0736 ± 0.0024	0.9645 ± 0.0444	0.0441 ± 0.0076	6.2 ± 0.8582	−0.003 ± 0.0030
GRU	0.0711 ± 0.0024	0.9736 ± 0.0444	0.0429 ± 0.0076	6.0 ± 0.8582	+0.002 ± 0.0030
D-GRU	0.0697 ± 0.0024	0.8788 ± 0.0444	0.0649 ± 0.0076	8.5 ± 0.8582	−0.006 ± 0.0030
Bi-GRU	0.0701 ± 0.0024	0.9771 ± 0.0444	0.0508 ± 0.0076	7.1 ± 0.8582	+0.001 ± 0.0030
LSTM + CNN + GRU	0.0739 ± 0.0024	0.9700 ± 0.0444	0.0434 ± 0.0076	6.3 ± 0.8582	−0.002 ± 0.0030
Bi(LSTM + GRU) with Attention	0.0684 ± 0.0024	0.9832 ± 0.0444	0.0414 ± 0.0076	5.8 ± 0.8582	+0.002 ± 0.0030
TFT	0.0709 ± 0.0024	0.8745 ± 0.0444	0.0449 ± 0.0076	6.4 ± 0.8582	−0.004 ± 0.0030
LSTM + XGBoost Ensemble	0.0753 ± 0.0024	0.9578 ± 0.0444	0.0473 ± 0.0076	6.7 ± 0.8582	−0.003 ± 0.0030

2.3. Data Quality Routine and Pre-processing

Hourly data was averaged into daily observations to improve the dataset's quality. This aggregation reduces noise, smoothens temporal variations, and helps the model generalize better by focusing on broader patterns rather than transient fluctuations. Additionally, any NaN values in the dataset were filled out using mean imputation, ensuring data continuity and completeness. Averaging data is particularly advantageous in machine learning workflows, as it simplifies computational complexity and improves model performance when dealing with long-term trends. Missing values in the dataset was rare and did not form long continuous gaps. To maintain temporal consistency without artificially altering patterns, mean imputation was applied. Given the limited and scattered nature of these missing entries, this strategy introduced negligible impact on temporal dependencies, while ensuring data continuity. Data normalization methods, including standardization and Min-Max normalization [32], are commonly used to prepare datasets for machine learning models [33]. In this study, the Min-MaxScaler [22] was employed to scale the selected features within a range of 0 to 1, ensuring consistency across variables. This approach minimises potential biases during model training and enhances the efficiency of the learning process by improving convergence. The Min-Max normalization technique follows a linear transformation, as expressed in Equation (1), and is widely recognized for its effectiveness in handling diverse feature scales [34].

$$x' = \frac{x - \min(max_n - min_{n)}}{max - min} + min_n \tag{1}$$

The dataset was split strictly in chronological order to avoid temporal leakage. The first 80% of the data (covering earlier years) was used for training, within which 20% was reserved for validation, while the most recent 20% was allocated for testing. This approach ensured that all evaluations were conducted on unseen future data, preserving the temporal integrity of the forecasting task. To mitigate overfitting, dropout layers were incorporated in each architecture, and training was monitored with early stopping. Validation was performed using a rolling-origin chronological strategy to avoid temporal leakage and to ensure robustness of the selected configurations. Several analyses were conducted to visualize the power generation trends over time. Fig. 5 illustrates the average power generation trends.

As illustrated in Fig. 5, annual average power generation revealed a

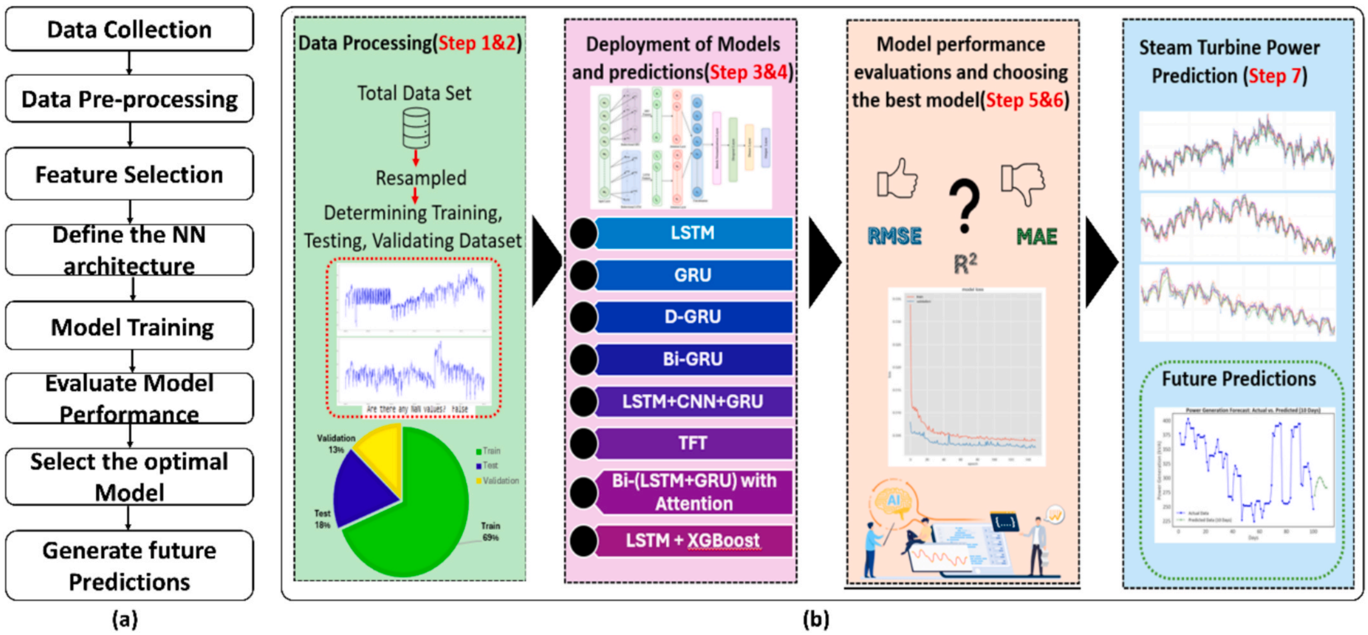

Fig. 3. Overview of the proposed work. (a) Workflow of the study and (b) Step-by step framework for time-series forecasting.

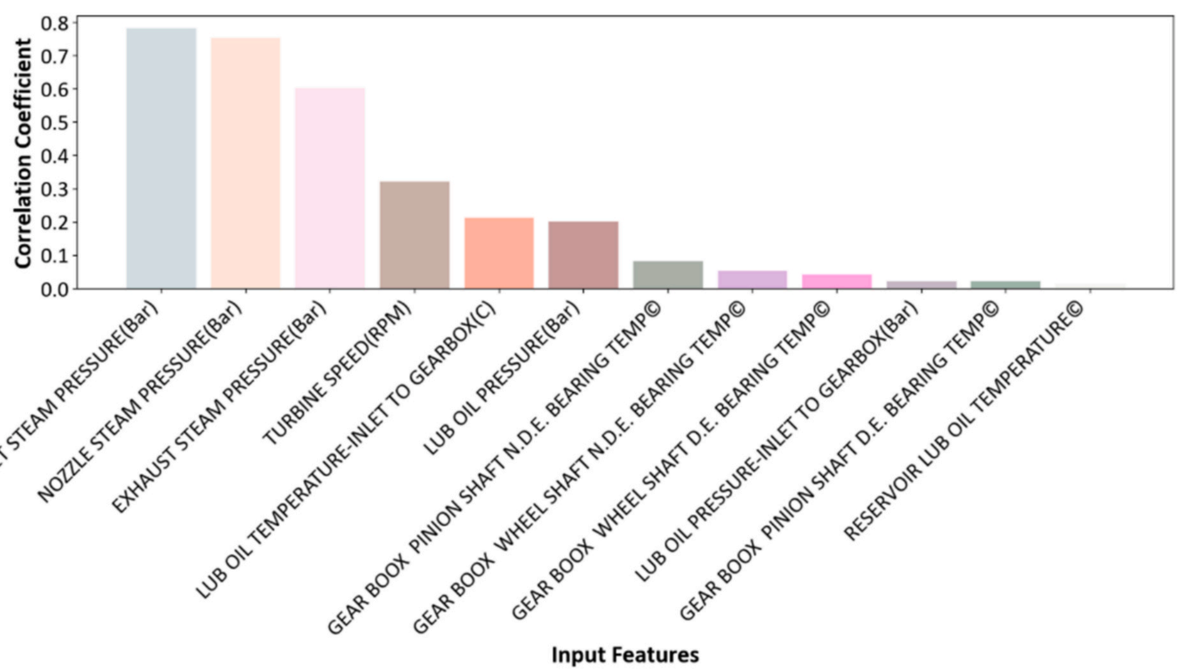

Fig. 4. Correlation Coefficients of Input Features with Power Generation Output.

consistent pattern with slight yearly variations. Monthly average power generation emphasizes seasonal variations across the eight years, while weekly patterns highlight how power generation varies by the day of the week. Daily average power generation trends over the entire dataset, provided granular insights into the variations across specific days. Steam turbine energy generation data for eight years are deposited to the repository as supplementary materials. Temporal windows [24] were incorporated during data pre-processing to structure the time series data into sequences of 100 prior observations rather than relying solely on timestamp data. This approach allows the model to focus on the relationships and dependencies between consecutive data points as capturing temporal patterns and long-term trends are essential for accurate predictions.

2.4. Data-driven machine learning models

This research aimed to develop a robust and accurate forecasting model for power generation in biomass-fuelled steam turbines. The baseline models evaluated in this study include LSTM, GRU, D-GRU, Bi-GRU, LSTM+CNN+GRU, Temporal Fusion Transformer (TFT), and LSTM+XGBoost, representing conventional recurrent networks, hybrid deep learning architectures, and transformer-based forecasting approaches. Noteworthy, that baseline forecasting approaches such as persistence and seasonal averages have already been consistently outperformed by deep learning models like LSTM and GRU [10,19,23,24,31]. In line with this evidence, the proposed Bi-(LSTM+GRU) with Attention model was benchmarked primarily against

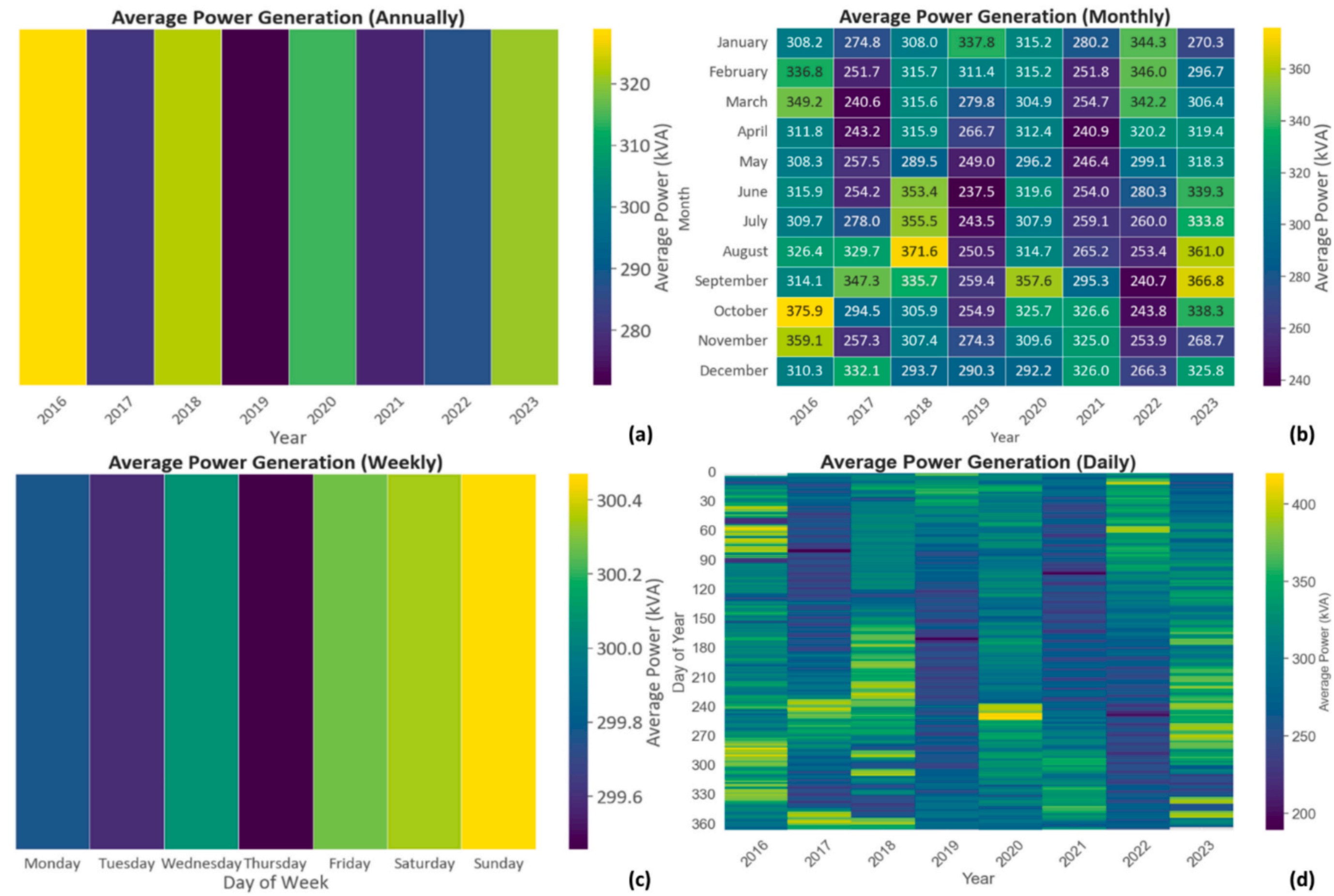

Fig. 5. Visualization of average power generation patterns at different temporal resolutions. (a) Annual average power generation, (b) Monthly average power generation for eight years, (c) Weekly average power generation, and (d) Daily average power generation.

advanced AI models, which provides a more rigorous and relevant comparison. The methodology involved iterative experimentation, with each model and evaluated its predictive accuracy and ability to handle the time-series nature of the dataset.

2.4.1. *Long Short-Term Memory*

The initial model was LSTM proposed by Hochreiter [17] and was particularly effective for time-series forecasting. As depicted in Fig. 6(a), the LSTM cell incorporates specialized gates, including the forget gate (f_t), input gate (i_t), and output gate (o_t), along with a control gate ($\widehat{c_t}$).

The LSTM cell processes the input X_t and previous hidden state h_{t-1} to calculate the updated cell state C_t and hidden state h_t , thereby capturing both short-term and long-term dependencies. The LSTM architecture in this study was designed with four sequential LSTM layers, each consisting of 50 units, and dropout regularization was incorporated at a rate of 20% to prevent overfitting. A Dense layer was used as the output layer to generate the final predictions. The training process incorporated the Adam optimisation with a learning rate of 0.001, employing MSE as the primary loss function and MAE as an additional evaluation metric. The model was trained for 150 epochs with a batch size of 32, using 20% of the training data for validation. After training, the model's performance was evaluated on test data using metrics such

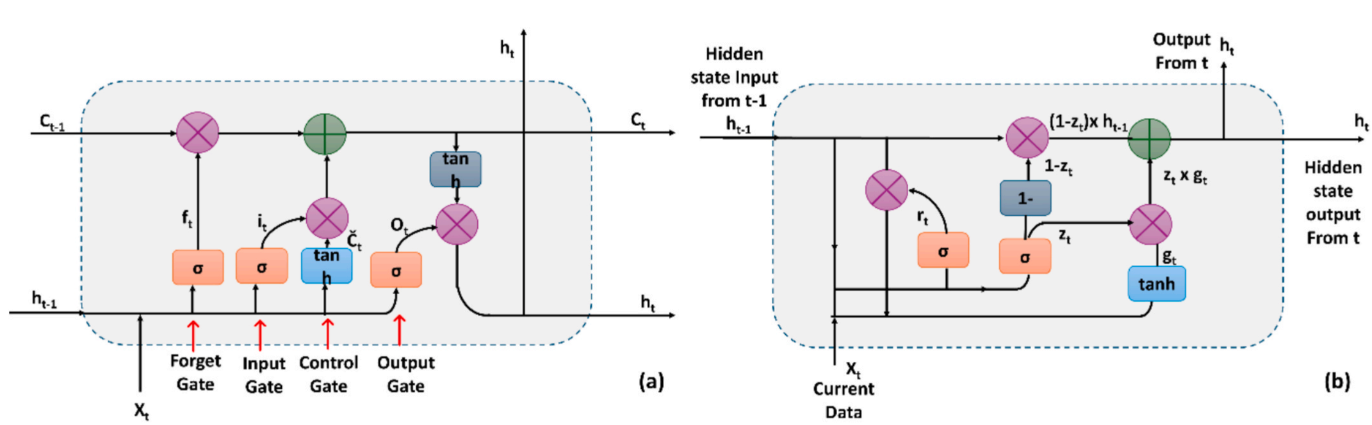

Fig. 6. Schematic diagram of LSTM and GRU cells. (a) LSTM cell and (b) GRU cell.

as RMSE, MSE, MAE, and R-squared (R2) to assess its predictive accuracy comprehensively.

2.4.2. Gated Recurrent Units

GRU introduced by Xiao et al.[35], was employed in this study for time-series forecasting due to their efficiency and simpler architecture compared to LSTM. As depicted in Fig. 6(b), the GRU cell incorporates specialized gates, including the reset gate (r_t) and update gate (z_t), which work together to regulate the flow of information within the network. The GRU model comprised of four sequential layers, each containing 50 units, with a dropout rate of 30% to mitigate overfitting. A Dense layer was added as the output layer to produce final predictions. The model training utilised the Adam optimiser with a learning rate 0.001, adopting Mean Squared Error (MSE) as the primary loss function and Mean Absolute Error (MAE) as a secondary evaluation metric. Training was conducted for 150 epochs with a batch size of 32, and 20% of the training dataset was reserved for validation. Evaluation metrics such as RMSE, MSE, MAE, and R-squared (R2) demonstrated the model's ability to effectively capture temporal dependencies, making it a reliable choice for forecasting power generation patterns.

2.4.3. Drop GRU

The Drop-GRU model, inspired by Mahjoub et al.[36], was adopted to improve temporal feature learning while addressing overfitting challenges in time-series data. The architecture consisted of four sequential GRU layers, each with 50 units and a 20% dropout rate, to effectively balance complexity and ensure generalization. This setup allowed the model to capture intricate temporal dependencies, making it well-suited for forecasting power generation trends.

2.4.4. Bi-GRU

The Bi-GRU model, inspired by advancements in RNN-based load forecasting, was experimented to address temporal dependencies in power generation data. As described by He et al.[37], Bi-GRU enhances the conventional GRU by processing data in both forward and backward directions, improving its ability to capture bidirectional relationships in time-series data. Zhang and Gu [38] and Jincheng [39] have highlighted the importance of temporal modelling in load forecasting for high-proportion renewable energy systems. The architecture of this experiment was designed with three stacked Bidirectional GRU layers, each incorporating 64 units. Batch normalization layers were added after each Bi-GRU block to ensure stable gradient flow and improved convergence during training. Dropout regularization of 20% was applied to prevent overfitting. Further, the Adam optimiser was used with a learning rate of 0.0001 for efficient parameter updates. The model was trained in over 150 epochs using 80% of the dataset, with 20% reserved for validation.

2.4.5. LSTM + CNN + GRU

The hybrid architecture combining LSTM, CNN and GRU was implemented to harness the individual strengths of these networks. LSTM layers effectively captured the temporal dependencies in the data. GRU layers provided computational efficiency, while maintaining robust feature extraction. Simultaneously, CNN layers extracted the spatial patterns. This combination addressed the dataset's spatial and temporal correlations, enhancing the model's prediction accuracy. The architecture included sequential layers, starting with an LSTM layer followed by GRU and CNN components. Dropout regularization was applied throughout the network to prevent overfitting. The Adam optimiser with 0.001 learning rate was used for training, ensuring efficient convergence. Each layer leveraged its unique ability to capture data complexities in terms of spatial features, long-term temporal patterns and short-term temporal patterns. The model was trained for 150 epochs with a batch size of 32, and validation was performed on 20% of the training dataset.

The hybrid structure was inspired by the works of Alharkan et al. [40], which successfully combined CNN and LSTM for solar power prediction, and Abumohsen et al.[41], who demonstrated the effectiveness of GRU models in load forecasting for power systems. Additionally, findings from Sajjad et al.[42] , which explored various hybrid deep learning models for time-series forecasting in energy applications, further reinforced the methodological choices. These studies collectively guided the development of the hybrid model by integrating the strengths of CNN, LSTM, and GRU architectures, ensuring robust and accurate predictive performance.

2.4.6. Temporal Fusion Transformer (TFT)

The TFT model was employed in this study to forecast power generation, leveraging its capability to handle multivariate time-series data and provide interpretable predictions. As outlined by Zheng et al. [43], the TFT model incorporates a multi-head attention mechanism to capture long-term dependencies and a variable selection network to prioritize key features from input data. The architecture was configured with six attention heads and a hidden size of 128, allowing the model to learn complex patterns in energy consumption data efficiently. Dropout regularization of 20%, was applied to prevent overfitting, and the Adam optimiser with a learning rate of 0.001, facilitated effective parameter updates. Training was conducted over 150 epochs, utilising 80% of the data for model fitting and reserving 20% for validation, ensuring robust performance evaluation.

2.4.7. LSTM + XGBoost

The hybrid model integrating LSTM and XGBoost implemented in this study leverages the strengths of both algorithms to improve predictive accuracy. LSTM was employed to capture sequential dependencies within the time-series data, while XGBoost addressed the residual errors, extracting non-linear relationships. It is a sophisticated ensemble learning framework [44]. This approach aligns with prior research by Frifra et al. [45] on storm prediction and Audace et al. [46] for photovoltaic power forecasting, highlighting the synergy of LSTM's temporal modelling with XGBoost's regression capabilities . The LSTM model was constructed using two layers, each with 50 units, followed by dropout regularization to mitigate overfitting. It was trained for 150 epochs with an Adam optimiser to minimise the mean squared error. Post-training, the extracted features from the LSTM were fed into an XGBoost regressor and optimised using parameters such as a learning rate of 0.1 and a maximum tree depth of 5. The combined architecture was validated using R2 and RMSE metrics to demonstrate enhanced prediction performance compared to standalone models.

2.4.8. Bi-(LSTM+GRU) with Attention

The proposed Bi-(LSTM+GRU) with Attention model represents a novel ensemble approach tailored explicitly for accurate and efficient energy load forecasting. Ensemble methods, which have gained popularity for their improved prediction accuracy compared to single models, combine the strengths of multiple base models to reduce generalization errors and improve robustness. Inspired by the foundational work of Hansen and Salamon[47], this ensemble model strategically integrates the bidirectional capabilities of LSTM networks, GRU's computational efficiency and an attention mechanism interpretability. By leveraging the collective decision-making power of its components, the model effectively addresses the unique challenges posed by highly volatile and complex time-series data, ensuring more reliable forecasts. The fundamental building blocks of the Bi-GRU and Bi-LSTM models are the GRU and LSTM cells, respectively, as depicted in Fig. 6. These root cells provide the computational foundation for the bidirectional architecture by effectively learning long-term dependencies while addressing vanishing gradient issues.

The BiLSTM component illustrated in Fig. 7(a) ensures the model captures both forward (S_f) and backward (S_b) temporal dependencies, which is a critical feature for sequential energy data where future and past states may interact intricately. BiGRU [48,49] illustrated in Fig. 7

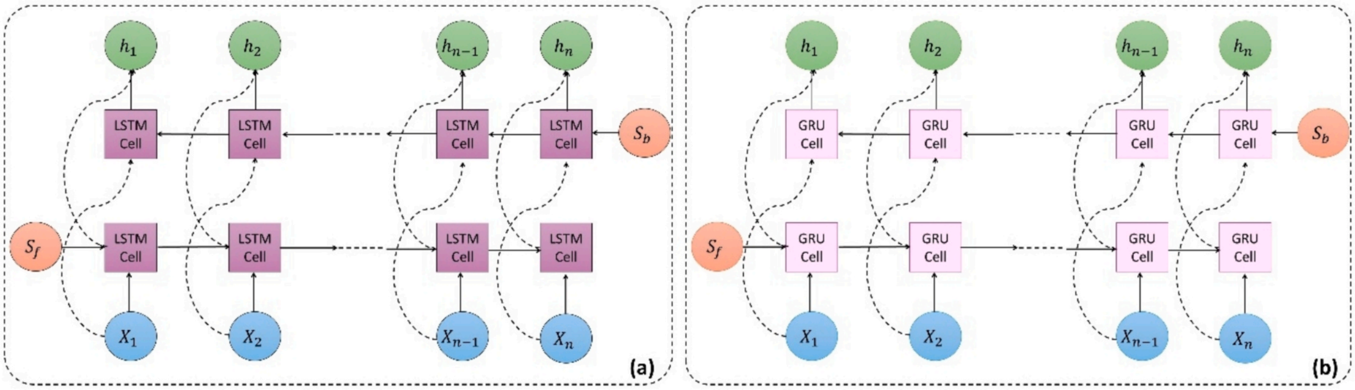

Fig. 7. Schematic diagram of Bi-LSTM and Bi-GRU. (a) Bi-LSTM cell and (b) Bi-GRU cell.

(b), on the other hand, complements this by reducing computational overhead while preserving the core functionality of capturing long-term dependencies, as demonstrated in prior research .The Bi-GRU and Bi-LSTM architectures are designed to capture both past and future dependencies in time-series data, making them highly effective for complex sequential tasks.

The input sequence X= $\{X_1, X_2, \ldots, X_n\}$ is processed by two parallel pathways, the forward and backward recurrent layers. In the forward GRU/LSTM, denoted as h_t^f , information flows from the initial temporal window X1 to the final step X_n , capturing past temporal dependencies , producing hidden states $h_1^f, h_2^f, h_3^f, \cdots\cdots..h_n^f$. This process captures temporal dependencies from past to present. Conversely, the backward GRU/LSTM, denoted as h_t^b , processes the input sequence in reverse, from Xn back to X1 , to encode future contextual information, yielding hidden states $h_1^b, h_2^b, h_3^b, \cdots\cdots..h_n^b$. This captures dependencies from future to past. At each temporal window t, the hidden states h_t^f , and h_t^b , from the forward and backward pathways are concatenated to form a unified hidden representation h_t. By unifying these two h_t^f , and h_t^b , representations, the model gains a holistic understanding of past and future temporal features, enhancing its ability to capture intricate patterns in sequential data. This operation is represented mathematically by equation (2) below.

$$h_t = \left[h_t^f; h_t^b\right] \tag{2}$$

The initial state S_f of the forward layer and the final state S_b of the backward layer serve as boundary conditions for the GRU/LSTM cells. The resulting output, y_t, at each temporal window, is computed based on these combined features. To compute , y_t, the unified hidden representation h_t is passed through a fully connected dense layer that applies a transformation using a weight matrix W_0 and a bias vector b_0, followed by an activation function σ, as given by the equation (3).

$$y_t = \sigma(W_0 h_t + b_0) \tag{3}$$

The attention mechanism dynamically prioritizes temporal windows that influence predictions most, enabling the model to allocate its focus where it is most needed, thus enhancing interpretability and robustness. The attention mechanism computes weights α_t for each temporal window t by evaluating the relevance of the hidden states h_t to the output context. This process highlights key features in the time series, ensuring that the model captures irregular fluctuations and peak intervals effectively. The attention mechanism begins by calculating a score e_t for each temporal window, which measures the relevance of the corresponding hidden state to the current context as shown in equation (4), where s is a context or query vector that guides the focus of attention, and *f()* is a scoring function.

$$e_t = f(h_t, s) \tag{4}$$

These scores are then normalized using the softmax function to produce attention weights α_t as given by equation (5).

$$\alpha_t = \frac{\exp(e_t)}{\sum_{t=1}^{n} \exp(e_t)} \tag{5}$$

Once the attention weights are obtained, they are used to compute a context vector c, which aggregates the most relevant features from the hidden states. The context vector is derived as a weighted sum of the hidden states as given in equation (6).

$$c = \sum_{t=1}^{n} \alpha_t h_t \tag{6}$$

The context vector c, derived from the attention mechanism, represents the most significant features from the temporal sequence, enabling the model to focus on critical temporal windows while discarding irrelevant information. Building on this foundational mechanism, the proposed Bi-(LSTM+GRU) with Attention model integrates the outputs of BiLSTM and GRU layers within a unified ensemble framework. This integration leverages attention-weighted BiLSTM outputs to refine features in the GRU layer, ensuring that the most relevant bidirectional patterns are captured and utilised for accurate predictions. Such as BiLSTM-Attention [50] and BiLSTM-GRU combinations [49,48,51,52] this model integrates these architectures into a unified framework. However, the proposed Bi-(LSTM+GRU) with Attention model goes beyond simple integration It innovates by allowing the BiLSTM and BiGRU branches, each equipped with an attention mechanism, to extract bidirectional temporal features independently. The attention-weighted outputs from both branches are then combined through a fusion layer to generate the final prediction. This architecture capitalizes on the ability of BiLSTM to capture intricate bidirectional patterns, At the same time, GRU mitigates overfitting and reduces computation costs, thereby achieving an optimal trade-off between accuracy and efficiency.

While Drop-GRU and Bi-GRU variants offered incremental improvements over standard GRU, their standalone performance remained lower than the proposed ensemble. Drop-GRU suffered from reduced ability to model long-term dependencies due to high dropout regularization, whereas Bi-GRU demonstrated the advantage of bidirectional processing by improving R2 and reducing RMSE. However, only when bidirectional features were fused with Bi-LSTM and refined through attention weighting did the architecture achieve consistently superior performance. Heatmap analysis of normalized attention weights over a 100-step lookback window is given in Fig. 8 This ablation confirms that the ensemble's additive design is critical to achieving robustness and accuracy in non-linear steam turbine forecasting. Further, The attention mechanism was analysed to verify its contextual relevance. Weight matrices were extracted and visualized, confirming that the attention

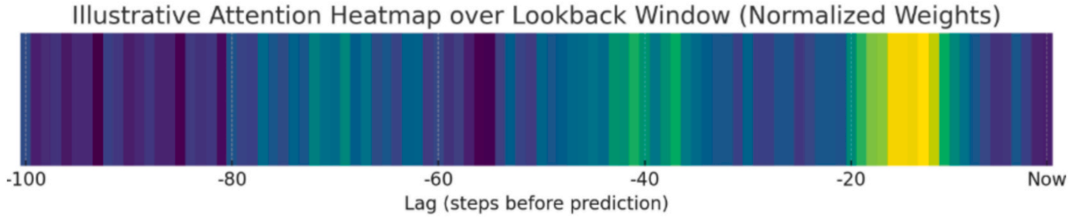

Fig. 8. Attention weight analysis. Heatmap of normalized attention weights over a 100-step lookback window for a representative prediction (right = most recent).

layers prioritized temporal windows with high variability, such as load spikes and sudden drops. This focus allowed the model to predict accurately even under uncertain load conditions. Although the inclusion of attention layers slightly increased computational complexity, they significantly enhanced interpretability by identifying the most critical time steps contributing to the forecast.

The novelty of the Bi-(LSTM+GRU) with Attention model lies in its ensemble structure, which seamlessly combines the interpretability of attention mechanisms with the complementary strengths of BiLSTM, BiGRU layers. Further, Bi-LSTM and Bi-GRU pathways independently compute their respective context vectors, capturing distinct bidirectional temporal dependencies. These context vectors are then concatenated to form a unified representation, integrating the prioritized features from both pathways. Unlike existing methods that often prioritize accuracy or computational efficiency, this approach delivers both by aligning feature prioritization with computational optimisation. Furthermore, dropout layers and a finely tuned Adam optimiser ensure generalization and model stability during training, even when faced with noisy or high-dimensional datasets. Extensive evaluations of the model, including metrics, consistently demonstrate superior performance compared to existing models. Hyperparameters were chosen via a grid search with rolling-origin chronological validation [53]. For the proposed Bi-(LSTM+GRU) with Attention model, we explored learning rate (1e-4, 5e-4, 1e-3), units per recurrent layer (32, 64, 128), dropout (0.1, 0.2, 0.3), batch size (16, 32, 64), and attention hidden size (32, 64, 128). Models were optimized with Adam optimizer, trained up to 200 epochs with early stopping (patience = 15), and selected by minimum average validation RMSE across folds. The final configuration was learning rate 0.001, 64 units per Bi-LSTM and Bi-GRU layer, dropout 0.2, attention hidden size 64, and batch size 32. Comparable grids were used for all baselines, which are listed in supplementary Table 2. This novel architecture not only addresses the inherent challenges of energy load forecasting and lays a foundation for future research in ensemble deep learning models tailored for energy systems and other time-sensitive applications. The ensemble is implemented using a stacked late-fusion approach. Bi-LSTM and Bi-GRU branches with attention are trained jointly, and their attention-weighted outputs are concatenated and fed into a fusion multilayer perceptron. This mechanism allows the model to learn adaptive weights for each branch, as opposed to fixed averaging. Ablation results given in Table 6 of supplementary data confirm the superiority of learned fusion over simple averaging.

To complement accuracy metrics, runtime, memory usage, and parameter counts were recorded for each model trained over 150 epochs. Table 3 of supplementary data illustrates that Bi-(LSTM+GRU) with Attention model required slightly longer runtime and higher memory (~105 minutes, 3.2 GB), while its inference time remained fast (<0.05s/batch), ensuring practicality for industrial deployment. Further, the accuracy gains observed sufficiently justify this moderate increase in computational cost.

In addition to training runtime, inference latency were evaluated. The Bi-(LSTM+GRU) with Attention model required <0.05s/batch on GPU and ~0.2s/batch on CPU, which demonstrates that real-time forecasting feasibility is retained. Thus, the proposed ensemble achieves a practical balance between accuracy and latency, ensuring its suitability for deployment in real-time industrial energy management.

Fig. 9 illustrates the proposed Bi-(LSTM+GRU) with Attention model, that improves accuracy of energy load forecasting. The model comprises sequential layers, including an input layer, bidirectional LSTM and GRU layers, attention mechanisms, batch normalization, concatenation, dropout, and a dense output layer, ensuring efficient extraction and integration of temporal features.

Paired t-tests were conducted on absolute errors [54] between the proposed and baseline models. For each test instance, error differences were computed, and mean differences, standard deviations, and t-statistics were derived. Results indicated significant improvements ($p < 0.05$) for the proposed model compared to all baselines, confirming the robustness of the performance gains.

2.5. Performance metrics

All models were trained on the normalized dataset, which was divided into training (64%), validation (16%), and testing (20%) subsets. The models were evaluated using key metrics, including Mean Absolute Error (MAE) [55], Root Mean Squared Error (RMSE) [55], and R-squared given below in equation (7), 8, and respectively. The selection of RMSE, MAE, and R-squared (R^2) as evaluation metrics was guided by their specific relevance to power prediction in dynamic energy systems. RMSE was chosen for its sensitivity to prediction errors, which is critical in power forecasting, where even slight deviations can significantly affect operational efficiency and energy distribution. MAE offers a straightforward and interpretable measure of average prediction error, ensuring that deviations, regardless of magnitude, are equally represented. R^2 evaluates the proportion of variance explained by the model, a vital factor for ensuring prediction reliability and consistency in fluctuating energy demand scenarios.

In addition to RMSE, MAE, and R^2, we evaluated MAPE and Bias (Mean Forecast Error) to quantify asymmetric forecasting errors. MAPE provides a scale-independent measure of error sensitivity, while Bias distinguishes systematic overestimation (positive) and underestimation

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{y}i - yi)^2} \tag{7}$$

$$MAE = \frac{\sum_{i=1}^{N}|\widehat{y}i - yi|}{\mathrm{N}} \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\widehat{y}i - yi)^2}{\sum_{i=1}^{N}(yi - \overline{y}i)^2} \tag{9}$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{\widehat{y}i - yi}{yi}\right| \tag{10}$$

$$MFE = \frac{1}{N}\sum_{i=1}^{N}(\widehat{y}i - yi) \tag{11}$$

3. Results and Discussion

A comprehensive experiment was conducted to forecast steam turbine power generation in a palm oil manufacturing facility, testing eight

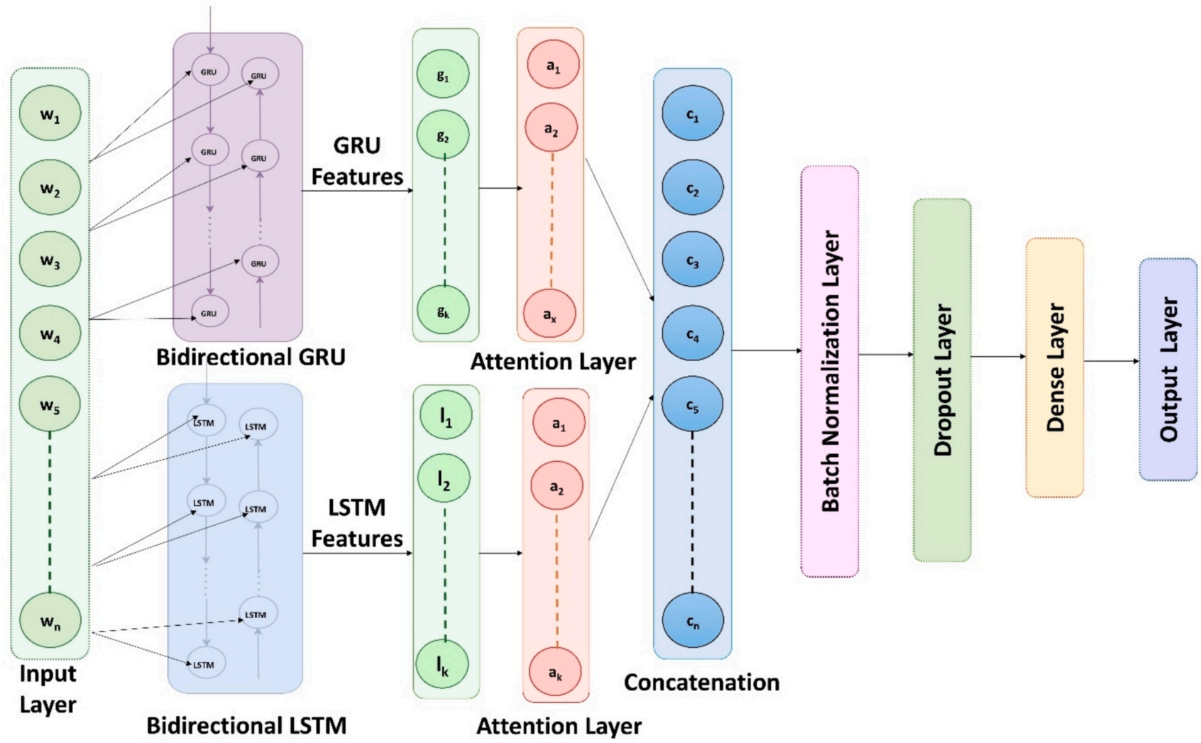

Fig. 9. Model architecture for steam turbine power prediction.

models, including the Bi-(LSTM+GRU) with Attention architecture, to address energy inefficiencies caused by fluctuating load demands. Each model's performance in predicting short and long-term trends, while laying the groundwork for optimised biomass resource management.

The Bi-(LSTM+GRU) with Attention model demonstrates superior performance compared to other benchmarks, including Bi-GRU, GRU, LSTM, and hybrid models such as LSTM+CNN+GRU and LSTM+XGBoost. It achieves consistently lower MAE across all epochs, reflecting faster convergence and enhanced prediction accuracy. The zoomed-in segments in Fig. 10 highlight significant performance gaps, particularly in later epochs, where the proposed model exhibits minimal error and robust generalization. The combination of dropout regularization, early stopping, and rolling-origin validation effectively prevented overfitting, as reflected by the stable convergence of training and validation losses across models.

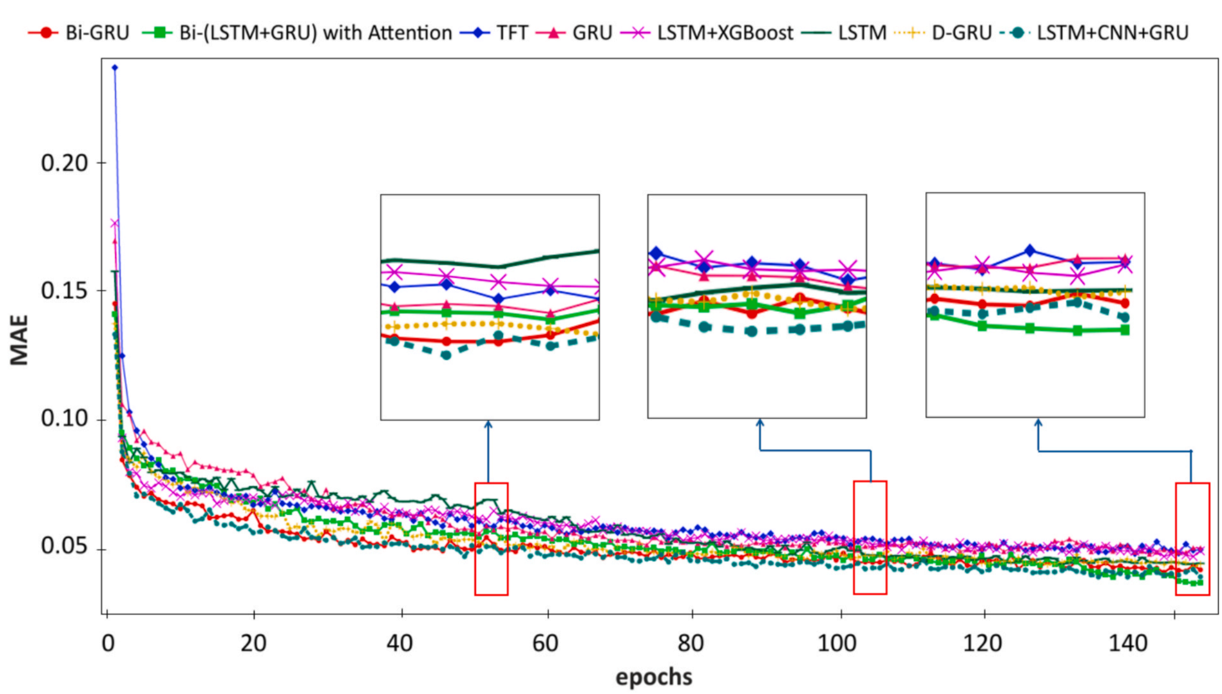

Fig. 10. MAE over epochs for models.

3.1. Analysis of model evaluation results

The radar plots illustrated in Fig. 11 provide a comparative analysis of key metrics, R-squared and RMSE. The Bi-(LSTM+GRU) with Attention model achieves the highest R-squared, demonstrating its capacity to effectively explain the variance in the dataset, and the lowest RMSE, indicating its precision in predictions. Competing models, including LSTM+CNN+GRU and LSTM+XGBoost, exhibit comparatively higher errors, reaffirming the proposed model's efficiency and adaptability in capturing the energy load data's complex and non-linear dynamics.

The evaluation metrics were reported with their standard deviations across folds. RMSE exhibited a mean of 0.0716 ± 0.0024, MAE a mean of 0.0475 ± 0.0076, and R^2 a mean of 0.9474 ± 0.0444. These values indicate that the performance is stable, with low variability, confirming the robustness of the proposed model. Results indicate that the proposed Bi-(LSTM+GRU) with Attention model achieves the lowest MAPE and near-zero Bias, thereby demonstrating robustness against asymmetric load fluctuations

Its bidirectional design leverages the strengths of LSTM and GRU layers, ensuring comprehensive temporal dependency capture, while the attention mechanism dynamically prioritizes critical features to enhance interpretability. Further, its performance metrics validate its ability to handle the dynamic and non-linear nature of power generation data, establishing it as a robust and efficient solution for energy forecasting.

3.2. Statistical Analysis of Prediction Results Using AI Models

3.2.1. Training Prediction Analysis

This section presents how predictions were performed for all the trained models, showcasing their ability to capture temporal dependencies and patterns in the data. The analysis highlights the prediction performance across training, validation, and testing datasets for each model.

Fig. 12 illustrates the training prediction results, showcasing the performance of various AI models in predicting steam turbine power generation over the initial temporal windows (101–allocated for training). The curve starts from the 101st step because the data preprocessing includes using a temporal window of 100 previous observations for each prediction. This sliding window approach ensures that the model learns temporal dependencies from the prior 100 temporal windows, making the first prediction possible only after sufficient data points (100) have been observed.

The predictions generated by the models are compared with actual power generation values to evaluate their ability to learn patterns from the training dataset. The Bi-(LSTM+GRU) with Attention model closely tracks the actual values throughout the training period, accurately capturing both the fluctuations and stable trends in the data. Other models show reasonable performance but with noticeable deviations during abrupt changes in the data.

The zoomed-in sections of Fig. 12 highlight critical time segments, emphasizing where the models deviate or align with the actual power generation. The results confirm the capability of the models to capture temporal dependencies with varying degrees of accuracy, depending on their architecture and underlying mechanisms. The Bi-(LSTM+GRU) with Attention model achieves enhanced precision and consistency in power prediction during training, leveraging its advanced architecture to effectively model complex temporal dependencies and non-linear trends. This analysis provides a detailed understanding of the predictive performance during the training phase, forming the basis for subsequent validation and testing evaluations.

3.2.2. Validation Prediction Analysis

The results presented in Fig. 13 show the comparison between the actual power generation data (kVA) and the predictions made by various validation models across temporal windows. The models evaluated include Bi-GRU, Bi-(LSTM+GRU) with Attention, TFT, GRU, LSTM+XGBoost, LSTM, D-GRU, and LSTM+GRU+CNN. Each model's predictions are shown alongside the actual data to evaluate their accuracy and ability to capture the underlying patterns in power generation.

Several critical regions, with sudden changes or fluctuations in power generation are highlighted and magnified to showcase superiority of the proposed approach. These regions are particularly challenging for prediction models, making them key areas for performance evaluation. In most temporal windows, models demonstrated a strong alignment with the actual data, indicating their ability to capture general trends effectively. Notably, models such as Bi-(LSTM+GRU) with Attention and

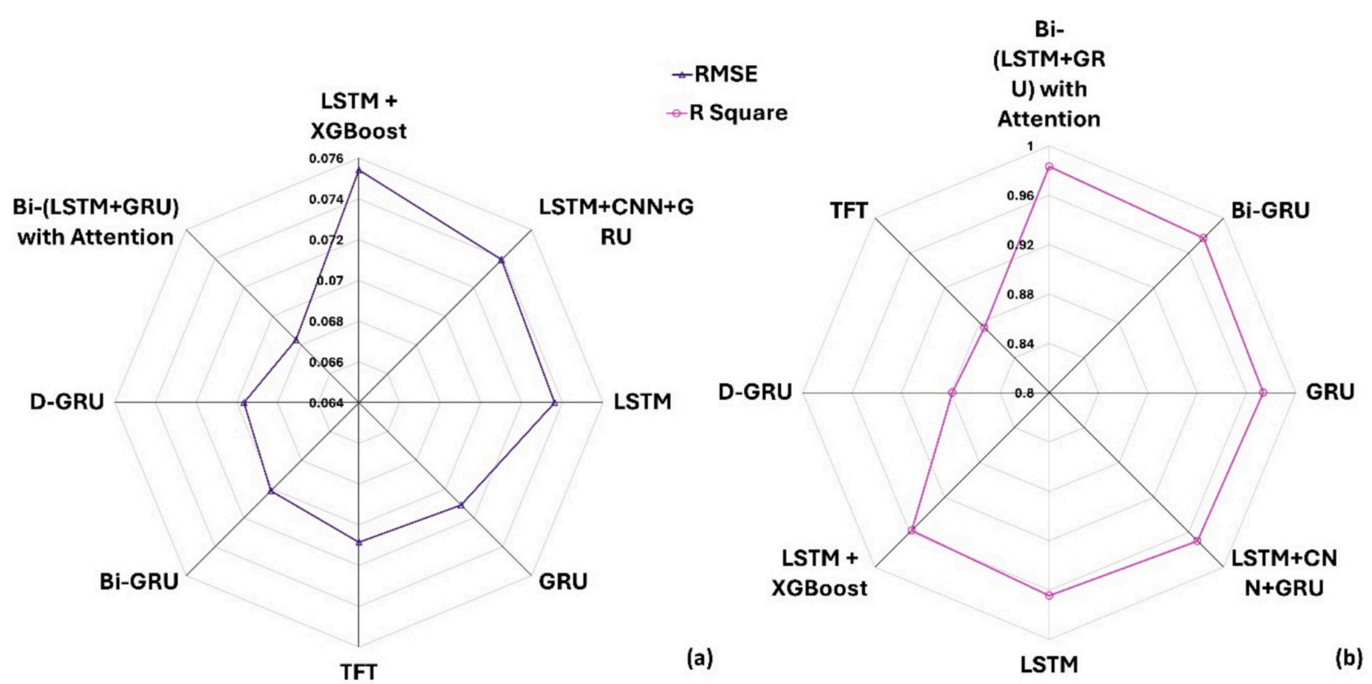

Fig. 11. Comparative analysis of key metrics. (a) RMSE evaluation and (b) R2 evaluation.

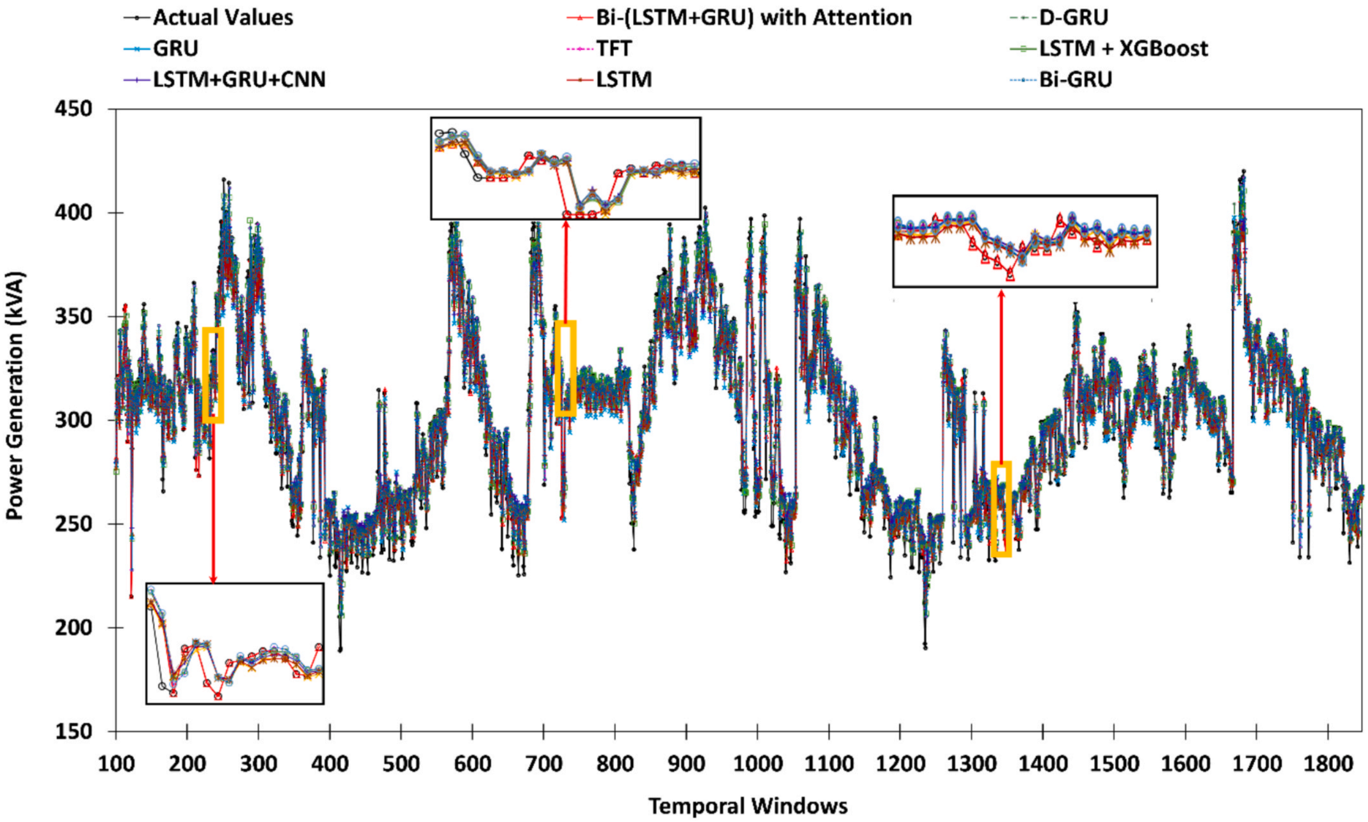

Fig. 12. Predicted and actual power generation for experimented models during model training.

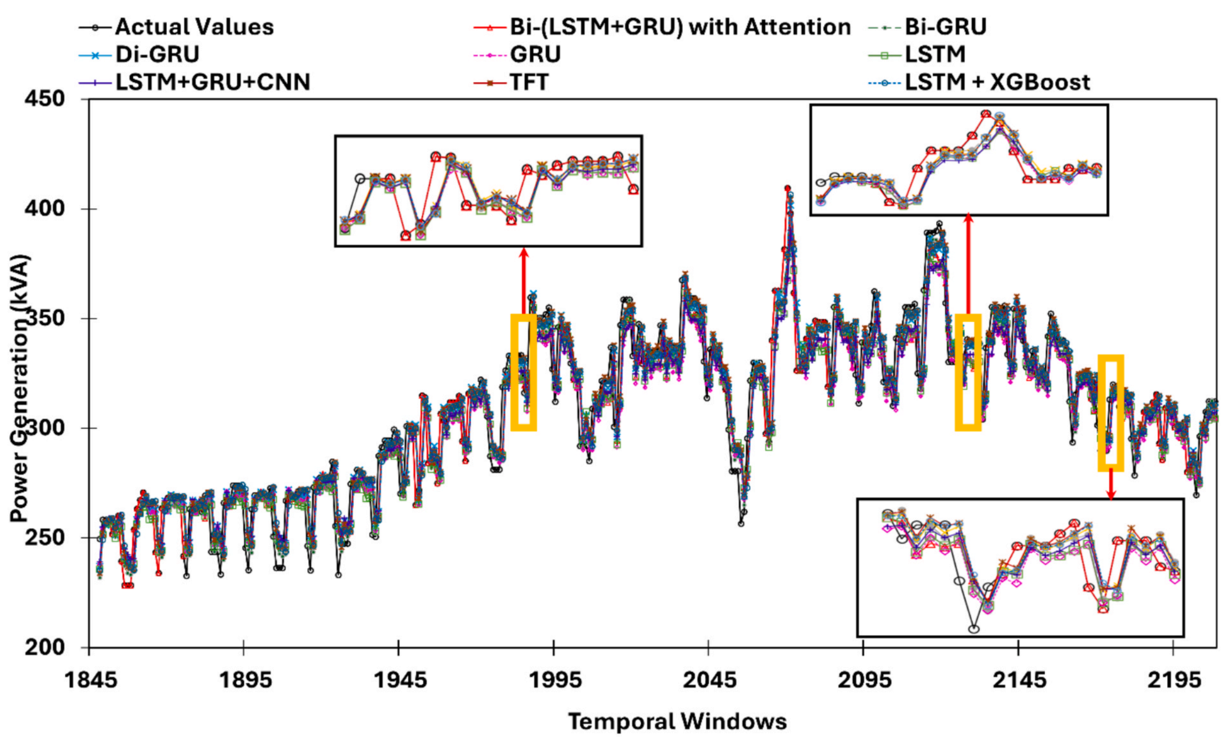

Fig. 13. Predicted and actual power generation for experimented models during model validation.

LSTM+GRU+CNN exhibit superior performance in these critical regions, accurately capturing sharp spikes and dips in power generation. Nevertheless, some other models i.e. GRU and LSTM+XGBoost, show deviations in these regions, highlighting their limitations in responding to abrupt changes in the data.

3.2.3. Testing Prediction Analysis

Fig. 14 presents the testing phase prediction analysis, which

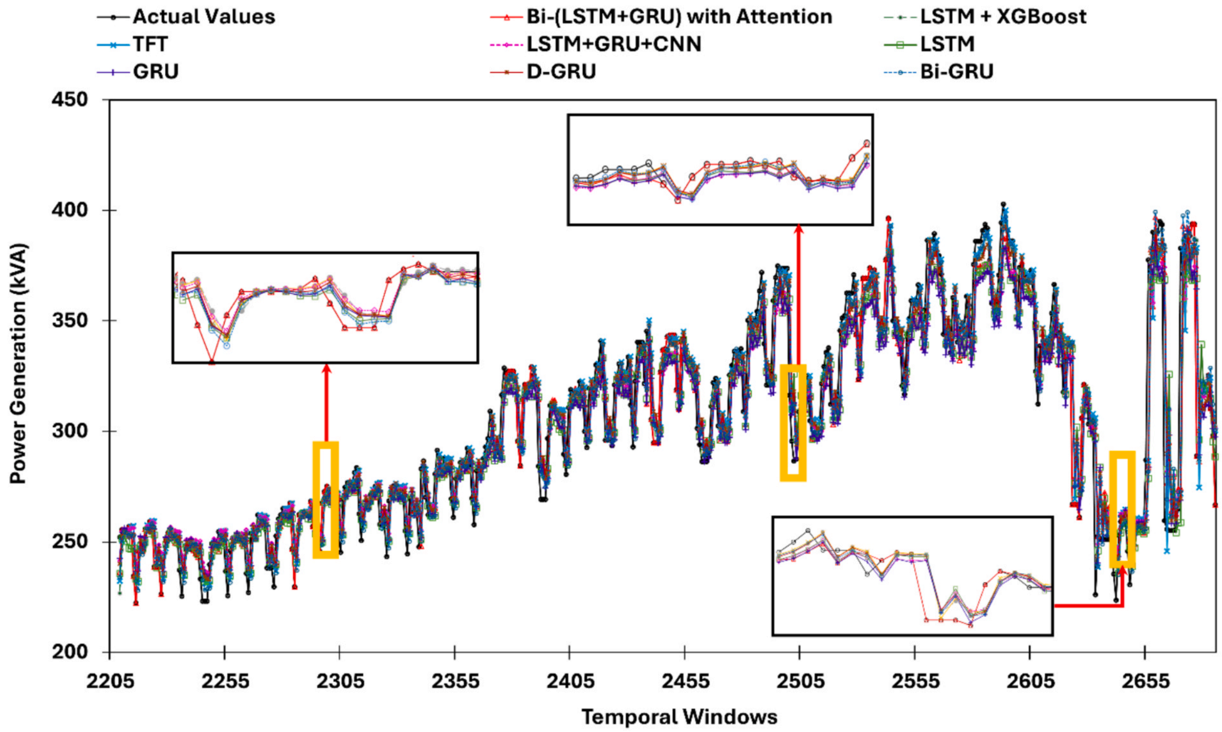

Fig. 14. Predicted and actual power generation for experimented models during model testing.

compares actual power generation data (KVA) with predictions made by various models across temporal windows 2200–2600. The models included in this analysis are Bi-GRU, Bi-(LSTM+GRU) with Attention, TFT, GRU, LSTM+XGBoost, LSTM, D-GRU, and LSTM+GRU+CNN. The selected temporal windows lie within the testing phase, which is critical for evaluating the generalization capability of the models on unseen data. Several regions of interest are highlighted with red rectangles and magnified to emphasize model performance under dynamic conditions.

In most cases, the models align closely with the actual power generation data, showcasing their ability to capture overall trends effectively during the testing phase. The highlighted regions, however, reveal differences in how well the models handle fluctuations and abrupt changes in power generation. For instance, Bi-(LSTM+GRU) with Attention and LSTM+GRU+CNN consistently provide predictions that closely aligns with actual values, even during sharp spikes and dips in power generation. In contrast, models such as GRU and LSTM+XGBoost exhibit slight deviations, particularly in regions with rapid transitions, suggesting limitations in their capacity to adapt to highly dynamic patterns. The zoomed-in areas in the figure also show the ability of specific hybrid models, such as Bi-GRU and D-GRU, to approximate the actual data during moderate fluctuations. However, minor discrepancies can still be observed in specific regions, indicating the inherent challenges of accurately modelling sudden changes in power generation. The zoomed-in areas provide critical insights into each model's strengths and weaknesses during the testing phase, particularly under scenarios involving rapid variations in the power generation trends.

3.2.4. Short- and Long-Term Forecasting Performance

In this section, forecasting is performed using the Bi-(LSTM+GRU) with Attention model, considering its superior performance during the model evaluation phase. The analysis focused on the model's ability to accurately predict power generation trends over short-term and long-term horizons, showcasing its robustness and reliability in handling varying temporal scales.

The comparison of predicted and actual power generation values, as depicted in Fig. 14, reveals a strong alignment along the 1:1 reference line, underscoring the model's high predictive accuracy. The scatter plot's colour gradient presented in Fig. 15 illustrates the distribution of predicted power levels, showcasing the model's ability to manage varying power generation scenarios effectively. The distribution of actual power values ranges between 225–400 kVA further reinforces the model's robustness, as the predictions remain consistent and unbiased across the entire dataset.

The concentration of data points near the diagonal line demonstrates the model's robustness in accurately capturing both low and high-power generation patterns with minimal deviation. While a few outliers indicate potential areas for further refinement, the overall alignment validates the model's capability to generalize effectively across diverse operational conditions. Hence, these results claim superior reliability and precision of proposed Bi-(LSTM+GRU) with Attention model's reliability and precision in forecasting power generation.

A detailed forecasting experiment was conducted to predict power generation trends incorporating Bi-(LSTM+GRU) with Attention model. The model's ability to forecast power generation over both short-term (10 days) and long-term (20 days) horizons was examined. The corresponding results are shown in Fig. 16 (a) and (b) respectively. The Bi-(LSTM+GRU) with Attention model consistently captured power generation patterns, closely aligning with actual data across the forecasted periods. The model effectively handled the variability in power generation, accurately predicting fluctuations and maintaining precision even over extended temporal scales. This reflects the robustness of the model in dealing with temporal dependencies and non-linear trends in the dataset. As illustrated in Fig. 16(a), the model accurately predicted power generation values with minimal deviations from the actual data for short-term forecasting. In long-term forecasting, shown in Fig. 16(b), the model maintained its performance, demonstrating its ability to generalize and provide reliable predictions over 20 days. This capability to accurately forecast power generation peaks and troughs is pivotal in addressing the inherent variability of power generation processes, particularly in dynamic scenarios such as those in the palm oil manufacturing sector.

The significance of this forecasting model extends beyond accurate predictions, directly addressing the challenges faced in the case study. This research bridges the gap between power generation variability and

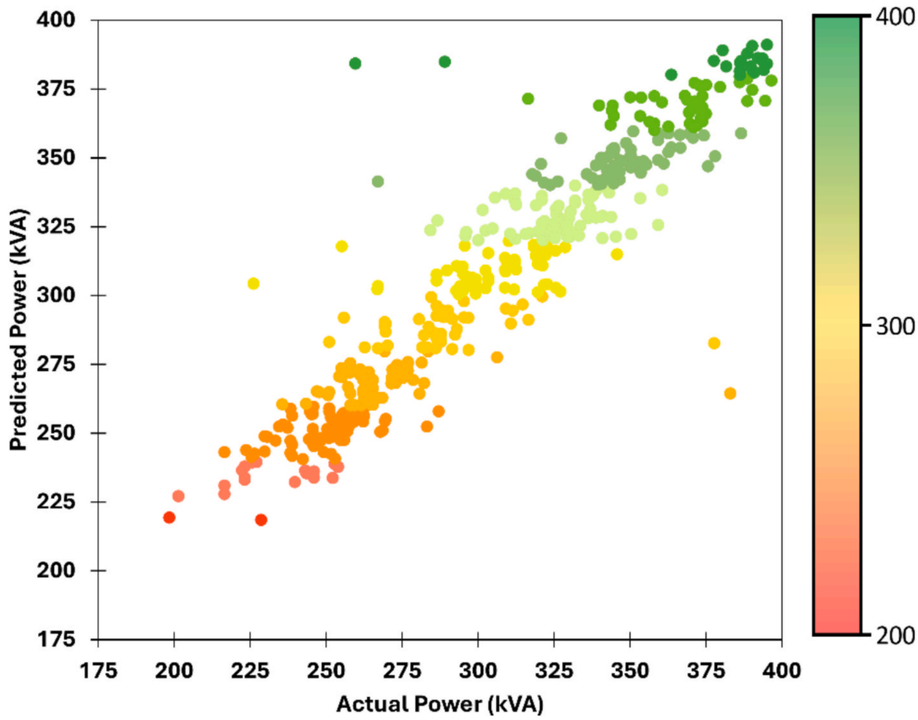

Fig. 15. Predicted vs. Actual power generation using proposed Bi (LSTM + GRU).

resource management by establishing a clear relationship between forecasted power generation and the availability of biomass fuel resources (palm fibres and palm shell). Further, this research bridges the gap between power generation variability and resource management as forecasted power generation values allow for a more precise estimation of fuel requirements, correlating generated power with Fresh Fruit Bunch (FFB) processing rates (tons/h). This insight enables operators to dynamically adjust fuel feed rates, ensuring an optimal balance between energy generation and biomass resource utilisation.

4. Conclusions and Perspectives

This research aimed to enhance the power generation efficiency in a palm oil manufacturing facility by developing an adaptive power forecasting model using a Bi-(LSTM+GRU) with Attention architecture. The primary objective was to address the challenges posed by fluctuating demands. These fluctuations often result in periods of excess energy production without corresponding load utilisation, leading to significant energy wastage. By accurately predicting power demands, the developed model improves energy utilisation and minimises reliance on grid power, which is both costly and resource intensive. This approach enhances operational efficiency and significantly reduces overall production costs, aligning with the dual objectives of enhancing sustainability and increasing profitability in a resource-intensive industry.

Building on the results of this research, the next phase focuses on optimising resource allocation, particularly in recommending optimal biomass, such as palm fruit fibres, shells, and husks, for steam turbine boilers. Integrating the forecasting model with resource management strategies enables precise recommendations on biomass utilisation, thereby mitigating the constraints of limited boiler feedstock. This step is crucial to achieving a balance between energy generation and resource availability, further contributing to the system's overall efficiency. The dataset, which includes parameters such as kVA, total unit energy, FFB (Fresh Fruit Bunch) production, and FFB per unit, forms the foundation for developing a resource recommendation system. The goal is to adjust biomass feed rates dynamically to align with forecasted energy demands, reducing both energy wastage and grid dependency. This approach addresses the core issue of inconsistent energy demand while ensuring that biomass resources are utilised effectively, paving the way for a more sustainable energy ecosystem. The Bi-(LSTM+GRU) with Attention model demonstrated superior performance in accurately predicting short-term and long-term power generation trends, enabling better planning and energy allocation. The approach supports cost reduction and higher profitability by reducing dependency on costly grid power and optimising steam turbine output, the approach supports cost reduction and higher profitability. Moreover, this work lays the groundwork for implementing a resource recommendation system, enabling optimised biomass utilisation and addressing boiler feedstock constraints. Future directions of this research include leveraging forecasting insights to recommend optimal biomass feed rates for boilers, ensuring efficient energy generation aligned with fluctuating demands. Real-time forecasting models capable of handling rapid changes in production schedules or energy requirements will also be explored.

Since the key objective of the study is to evaluate the feasibility of applying the proposed model for an industrial steam turbine context, the obtained results remarkably confirmed the suitability of proposed Bi-(LSTM+GRU) with Attention model for steam turbine power prediction. Hence, as future extensions we plan to test across different plants or industrial sites to confirm generalizability.

Although the proposed Bi-(LSTM+GRU) with Attention framework demonstrated high predictive accuracy and robustness, this study has a few minor limitations. As the analysis was conducted using data from a single industrial plant, some operational conditions specific to other facilities may not be fully represented. Nevertheless, the dataset's eight-year duration and comprehensive validation ensure that the results remain reliable and representative. Future work will focus on extending the study across multiple sites to further confirm the model's general applicability and scalability.

From a sustainability perspective, optimising biomass utilisation and reducing grid dependency contribute to environmental goals such as carbon footprint reduction and renewable energy integration. The insights from this study pave the way for advanced, data-driven solutions that align with the principles of sustainable development while driving economic value for palm oil production facilities. By addressing

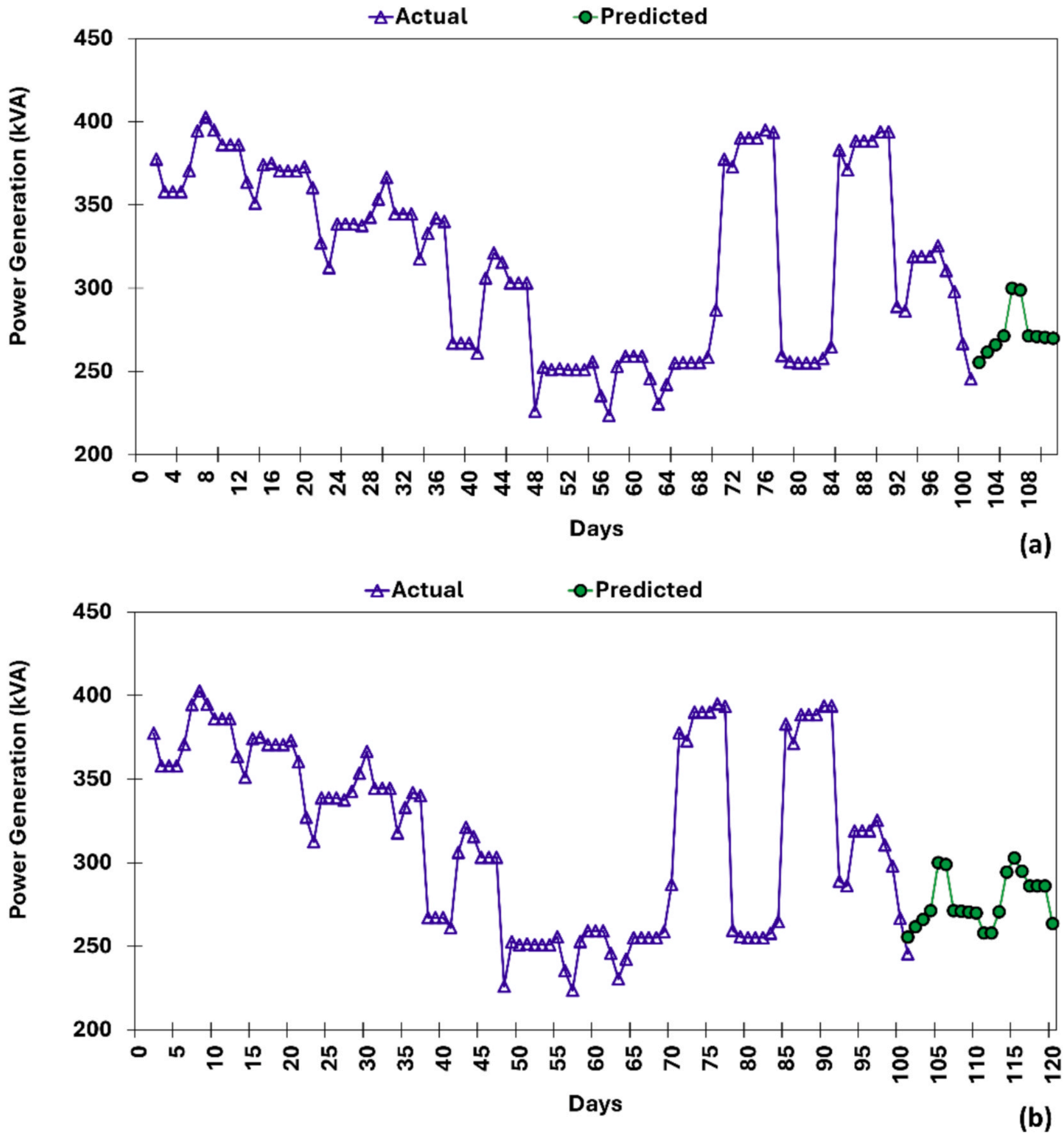

Fig. 16. Actual vs. predicted power generation. (a) Power generation forecast for 10-days and (b) Power generation forecast for 20-days.

inefficiencies and planning for future innovations, this research provides a solid foundation for adaptive energy management and resource optimisation in the palm oil industry.

The proposed Bi-(LSTM+GRU) with Attention model achieved a 25% improvement in prediction accuracy compared to benchmark models, significantly enhancing the reliability of steam turbine power forecasting in palm oil manufacturing. By accurately predicting power demand, the model provides a foundation for reducing grid dependency and improving biomass utilization efficiency. This contributes directly to operational cost reduction and sustainability goals. As future work, we plan to extend this framework towards joint biomass resource allocation, thereby quantifying the trade-off between fuel availability and energy generation efficiency.

CRediT authorship contribution statement

Himaya Perera: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Shalitha Jayasekara:** Methodology, Formal analysis, Conceptualization. **Ruchire Eranga Wijesinghe:** Writing – review & editing, Supervision, Funding acquisition. **Bhagya Nathali Silva:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Formal analysis. **Honnyong Cha:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-00562777) and partially funded by Sri Lanka Institute of Information Technology, Sri Lanka, under Grant PVC(R&I)/RG/2025/15 and Grant PVC(R&I)/RG/2025/28.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.enconman.2026.121464.

Data availability

Details are given in the manuscript

References

[1] N.-A. Mohd-Lair, J. F.-M. Lee, W. Y. Liew, and R. Sarbatly, “Simulation based process improvement in palm oil industry: A case study at palm kernel department,” in 2012 IEEE Business, Engineering & Industrial Applications Colloquium (BEIAC), IEEE, 2012, pp. 12–15. Accessed: Nov. 02, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6226037/?casa_token=0hAOeGOpBXEAAAAA:tSpbjmzUjaa6MHsyybzJ1fNq-dpIYnvhfVafPuPBE8GtagBlZRmM_On-KkK6PWtQ80sDFr1TK8iRzKA.

[2] Husain Z, Zainac Z, Abdullah Z. Briquetting of palm fibre and shell from the processing of palm nuts to palm oil. Biomass Bioenergy 2002;22(6):505–9.

[3] Mystakidis A, Koukaras P, Tsalikidis N, Ioannidis D, Tjortjis C. Energy Forecasting: A Comprehensive Review of Techniques and Technologies. Energies 2024;17(7):1662.

[4] Nikpey H, Assadi M, Breuhaus P. Development of an optimized artificial neural network model for combined heat and power micro gas turbines. Appl Energy 2013;108:137–48.

[5] Obuka Nnaemeka SP, Onyechi Pius C, Okoli Ndubuisi C. “Palm oil biomass waste a renewable energy resource for power generation,” Saudi. J Eng Technol 2018;3:680–91.

[6] Chung J, Jang B. Accurate prediction of electricity consumption using a hybrid CNN-LSTM model based on multivariable data. PLoS One 2022;17(11):e0278071.

[7] Cakiroglu C, Demir S, Ozdemir MH, Aylak BL, Sariisik G, Abualigah L. Data-driven interpretable ensemble learning methods for the prediction of wind turbine power incorporating SHAP analysis. Expert Syst Appl 2024;237:121464.

[8] Nasution MA, Herawan T, Rivani M. Analysis of palm biomass as electricity from palm oil mills in North Sumatera. Energy Procedia 2014;47:166–72.

[9] Wu Q, Qiang TC, Zeng G, Zhang H, Huang Y, Wang Y. Sustainable and renewable energy from biomass wastes in palm oil industry: A case study in Malaysia. Int. J. Hydrog. Energy 2017;42(37):23871–7.

[10] Pasandideh M, Taylor M, Tito SR, Atkins M, Apperley M. Predicting Steam Turbine Power Generation: A Comparison of Long Short-Term Memory and Willans Line Model. Energies 2024;17(2):352.

[11] M. Khan, B. N. Silva, O. Khattab, B. Alothman, and C. Joumaa, “A transfer reinforcement learning framework for smart home energy management systems,” IEEE Sens. J., vol. 23, no. 4, Art. no. 4, 2022.

[12] Silva BN, Khan M, Wijesinghe RE, Wijenayake U. Meta-heuristic optimization based cost efficient demand-side management for sustainable smart communities. Energy Build 2024;303:113599.

[13] Sun L, Smith R. Performance Modeling of New and Existing Steam Turbines. Ind Eng Chem Res Feb. 2015;54(6):1908–15. https://doi.org/10.1021/ie5032309.

[14] Al-Shami M, Mohamed O, Abu Elhaija W. Energy-Efficient Control of a Gas Turbine Power Generation System. Designs 2023;7(4):85.

[15] A. Wankhede and V. Ghate, “Prediction of power of a combine gas and steam turbine using artificial neural network,” in 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE, 2018, pp. 1103–1108. Accessed: Nov. 12, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9012143/?casa_token=crqzJB9E73IAAAAA:Nr3eGG4HYNihRSa47uHX4AInjolXa7phhQkVYUg7BKi1GtDrGf2qgUpy1yrhRqSxRrew1jKr-mWk5Mo.

[16] Mohamed O, Wang J, Khalil A, Limhabrash M. Predictive control strategy of a gas turbine for improvement of combined cycle power plant dynamic performance and efficiency. Springerplus Dec. 2016;5(1):980. https://doi.org/10.1186/s40064-016-2679-2.

[17] S. Hochreiter, “Long Short-term Memory,” Neural Comput. MIT-Press, 1997, Accessed: Nov. 04, 2024. [Online]. Available: https://sophieeunajang.wordpress.com/wp-content/uploads/2020/10/lstm.pdf.

[18] Talaat M, Gobran MH, Wasfi M. A hybrid model of an artificial neural network with thermodynamic model for system diagnosis of electrical power plant gas turbine. Eng Appl Artif Intell 2018;68:222–35.

[19] Fakir K, Ennawaoui C, El Mouden M. Deep Learning Algorithms to Predict Output Electrical Power of an Industrial Steam Turbine. Appl. Syst. Innov. 2022;5(6):123.

[20] Gupta RK, Singh RC. Optimizing high-speed rotating shaft vibration control: Experimental investigation of squeeze film dampers and a comparative analysis using Artificial Neural Networks (ANN) and Response Surface Methodology (RSM). Expert Syst Appl 2024;249:123800.

[21] Fu Y, Li Z, Zhang H, Xu P. Using support vector machine to predict next day electricity load of public buildings with sub-metering devices. Procedia Eng 2015;121:1016–22.

[22] Abdoos AA, Abdoos H, Kazemitabar J, Mobashsher MM, Khaloo H. An intelligent hybrid method based on Monte Carlo simulation for short-term probabilistic wind power prediction. Energy 2023;278:127914.

[23] Sun L, Liu T, Xie Y, Zhang D, Xia X. Real-time power prediction approach for turbine using deep learning techniques. Energy 2021;233:121130.

[24] Karaman ÖA. Prediction of wind power with machine learning models. Appl Sci 2023;13(20):11455.

[25] Dou Y, Tan S, Xie D. Comparison of machine learning and statistical methods in the field of renewable energy power generation forecasting: a mini review. Front Energy Res 2023;11:1218603.

[26] Kaboli SHA, Selvaraj J, Rahim NA. Long-term electric energy consumption forecasting via artificial cooperative search algorithm. Energy 2016;115:857–71.

[27] Ouyang T, He Y, Li H, Sun Z, Baek S. Modeling and forecasting short-term power load with copula model and deep belief network. IEEE Trans. Emerg. Top. Comput. Intell. 2019;3(2):127–36.

[28] K. Chandrarathna, A. Edalati, and A. F. tabar, “Forecasting Short-term load using Econometrics time series model with T-student Distribution,” Sep. 28, 2020, arXiv: arXiv:2009.13595. doi: 10.48550/arXiv.2009.13595.

[29] Chen J-F, Do QH, Nguyen TVA, Doan TTH. Forecasting monthly electricity demands by wavelet neuro-fuzzy system optimized by heuristic algorithms. Information 2018;9(3):51.

[30] Yang Y, Chen Y, Wang Y, Li C, Li L. Modelling a combined method based on ANFIS and neural network improved by DE algorithm: A case study for short-term electricity demand forecasting. Appl Soft Comput 2016;49:663–75.

[31] Agga A, Abbou A, Labbadi M, El Houm Y, Ali IHO. CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. Electr Power Syst Res 2022;208:107908.

[32] Ali PJM. Investigating the Impact of min-max data normalization on the regression performance of K-nearest neighbor with different similarity measurements. ARO-Sci. J. Koya Univ. 2022;10(1):85–91.

[33] Aggarwal CC. An Introduction to Data Mining. In: Data Mining. Cham: Springer International Publishing; 2015. p. 1–26. https://doi.org/10.1007/978-3-319-14142-8_1.

[34] Punyani P, Gupta R, Kumar A. A multimodal biometric system using match score and decision level fusion. Int J Inf Technol Mar. 2022;14(2):725–30. https://doi.org/10.1007/s41870-021-00843-3.

[35] Xiao Y, Zou C, Chi H, Fang R. Boosted GRU model for short-term forecasting of wind power with feature-weighted principal component analysis. Energy 2023;267:126503.

[36] Mahjoub S, Chrifi-Alaoui L, Marhic B, Delahoche L. Predicting energy consumption using LSTM, multi-layer GRU and drop-GRU neural networks. Sensors 2022;22(11):4062.

[37] H. He, H. Wang, H. Ma, X. Liu, Y. Jia, and G. Gong, “Research on short-term power load forecasting based on Bi-GRU,” in Journal of Physics: Conference Series, IOP Publishing, 2020, p. 012017. Accessed: Dec. 12, 2024. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1639/1/012017/meta.

[38] Zhang T, Gu J. Markov short-term load forecasting method for power system with high proportion renewable energy. Power Syst. Technol. 2018;42(4):1071–8.

[39] Jincheng F. Research on short-term power load forecasting model based on deep learning. China Chengdu China: Univ. Electron. Sci. Technol; 2020.

[40] Alharkan H, Habib S, Islam M. Solar power prediction using dual stream CNN-LSTM architecture. Sensors 2023;23(2):945.

[41] Abumohsen M, Owda AY, Owda M. Electrical load forecasting using LSTM, GRU, and RNN algorithms. Energies 2023;16(5):2283.

[42] Sajjad M, et al. A novel CNN-GRU-based hybrid approach for short-term residential load forecasting. IEEE Access 2020;8:143759–68.

[43] Zheng P, Zhou H, Liu J, Nakanishi Y. Interpretable building energy consumption forecasting using spectral clustering algorithm and temporal fusion transformers architecture. Appl Energy 2023;349:121607.

[44] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[45] Frifra A, Maanan M, Maanan M, Rhinane H. Harnessing LSTM and XGBoost algorithms for storm prediction. Sci Rep 2024;14(1):11381.

[46] D. K. B. Audace, A. R. Gilles, and A. B. Macaire, “LSTM and XGBoost Models for 24-hour Ahead Forecast of PV Power from Direct Irradiation”, Accessed: Dec. 14, 2024. [Online]. Available: https://www.researchgate.net/profile/Kossoko-Didavi/publication/374163018_LSTM_and_XGBoost_models_for_24-hour_photovoltaic_power_forecasting_from_direct_irradiation_data/links/657721a5ea5f7f02055f7556/LSTM-and-XGBoost-models-for-24-hour-photovoltaic-power-forecasting-from-direct-irradiation-data.pdf.

[47] Hansen LK, Salamon P. Neural network ensembles. IEEE Trans Pattern Anal Mach Intell 1990;12(10):993–1001.

[48] M. A. M. Fargalla, W. Yan, and T. Wu, “Attention-Based Bi-Directional Gated Recurrent Unit (Bi-GRU) for Sequence Shale Gas Production Forecasting,” in International Petroleum Technology Conference, IPTC, 2024, p. D021S040R001. Accessed: Dec. 14, 2024. [Online]. Available: https://onepetro.org/IPTCONF/proceedings-abstract/24IPTC/2-24IPTC/542801.

[49] Jung S, Moon J, Park S, Hwang E. An attention-based multilayer GRU model for multistep-ahead short-term load forecasting. Sensors 2021;21(5):1639.

[50] J. Du, Y. Cheng, Q. Zhou, J. Zhang, X. Zhang, and G. Li, “Power load forecasting using BiLSTM-attention,” in IOP Conference Series: Earth and Environmental Science, IOP Publishing, 2020, p. 032115. Accessed: Dec. 14, 2024. [Online]. Available: https://iopscience.iop.org/article/10.1088/1755-1315/440/3/032115/meta.

[51] M. E. Karim and S. Ahmed, “A deep learning-based approach for stock price prediction using bidirectional gated recurrent unit and bidirectional long short term memory model,” in 2021 2nd Global Conference for Advancement in

Technology (GCAT), IEEE, 2021, pp. 1–8. Accessed: Dec. 14, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9587895/.
[52] Gupta V. Bi-LSTM GRU-based deep learning architecture for export trade forecasting. Int J Comput Appl Technol 2024;74(3):235–45. https://doi.org/10.1504/IJCAT.2024.141933.
[53] Zhao Y, Zhang W, Liu X. Grid search with a weighted error function: Hyper-parameter optimization for financial time series forecasting. Appl Soft Comput 2024;154:111362.
[54] Yusop H, Yeng FF, Jumadi A, Mahadi S, Ali MN, Johari N. The effectiveness of excellence camp: A study on paired sample. Procedia Econ. Finance 2015;31: 453–61.
[55] Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. Int J Forecast 2006;22(4):679–88.