Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi
Vol.10 No.2 July 2025, P-ISSN : 2502-3470, E-ISSN : 2581-0367

Optimized Hybrid CNN-Residual BiLSTM with Adaptive Prediction System for Enhanced Gas Turbine Performance Forecasting

Andika Pratama[1*], Chastine Fatichah[2]

[1]*Master's Program in Technology Management, School of Interdisciplinary Management and Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*

[2]*Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*

[1]6032231200@student.its.ac.id (*)

[2]chastine@if.its.ac.id

Received: 2025-05-11; Accepted: 2025-07-22; Published: 2025-07-31

Abstract— Accurately forecasting critical performance parameters, such as Compressor Discharge Pressure (PCD), in gas turbines is a strategic imperative for ensuring operational reliability and energy efficiency, particularly in vital facilities like Central Processing Plants (CPPs). However, achieving reliable forecasts presents significant analytical challenges due to the complex multivariate, non-linear, and noisy nature of industrial sensor data, compounded by dynamic operational loads. This study introduces and validates an integrated analytical framework centered on a systematically optimized Hybrid Convolutional Neural Network-Residual Bi-Directional Long Short-Term Memory (CNN-Residual BiLSTM) architecture. This hybrid design synergistically leverages CNN layers for multi-scale temporal pattern extraction and Residual BiLSTM blocks for robust long-range dependency modelling, enhanced by residual connections for training stability. The framework emphasizes rigorous data pre-processing and the selection of a comprehensive feature set, incorporating thermodynamic, electrical, and operational control signals to provide a holistic view of the turbine's state. Automated hyperparameter optimization via the Optuna framework is employed to maximize the model's predictive potential. Empirical validation demonstrates that the optimized configuration's performance is superior to that of baseline models (RMSE = 0.0611, MAE = 0.0298, R^2 = 0.9601), confirming the framework's contribution to advancing data-driven performance diagnostics and predictive maintenance (PdM) strategies for gas turbines.

Keywords— Compressor Discharge Pressure (PCD); Deep Learning; Gas Turbine; Hybrid Model; Hyperparameter Optimization; Predictive Maintenance; Residual Bi-LSTM; Time Series Forecasting.

I. INTRODUCTION

Gas turbines serve as critical power generation assets in essential sectors, including energy production and oil and gas processing, underpinning the operational stability of facilities such as Central Processing Plants (CPPs) [1]. Ensuring high operational reliability and energy efficiency for these turbines is a strategic imperative, as it directly influences economic performance and the continuity of critical supplies. Consequently, sophisticated condition monitoring and predictive maintenance (PdM) frameworks are increasingly vital for proactive asset management [2][3]. Within the complex interplay of operational parameters, Compressor Discharge Pressure (PCD) stands out as a particularly sensitive indicator, reflecting the turbine's thermodynamic efficiency, proximity to operational limits like surge, combustion stability, and overall component health [1]. Therefore, the analytical capability to accurately forecast PCD dynamics offers substantial leverage for optimizing real-time operational decisions and implementing more effective, condition-informed maintenance strategies [4][5].

The analytical hurdles in generating accurate PCD forecasts are twofold, stemming from challenges inherent to both the operational data and the predictive models themselves. On one hand, the data is intrinsically complex; industrial sensor streams from gas turbines are high-dimensional and multivariate, characterized by pronounced non-linear relationships, significant measurement noise, and non-stationarity driven by dynamic load profiles [6]. On the other hand, these demanding data characteristics render conventional statistical models inadequate, as they often fail to capture such intricate temporal dependencies and non-linearities [7]. While standard deep learning architectures like Long Short-Term Memory (LSTM) networks [8] offer improvements, they can fall short of the required precision when confronted with such complex industrial dynamics without substantial architectural and parametric refinement [9][10].

Advanced deep learning techniques, particularly recurrent architectures such as Bidirectional LSTM (Bi-LSTM) and hybrid models combining Convolutional Neural Networks (CNNs) with LSTMs [11]–[13], alongside architectural enhancements like residual connections [14]–[16], represent the state-of-the-art for modeling complex sequential data. However, the mere application of these advanced architectures often fails to realize their full potential, with many studies still facing challenges in achieving high-precision forecasts on real-world industrial datasets [18]. Achieving consistently high forecasting accuracy is critically sensitive to the model's specific configuration—its hyperparameters [17]. A significant gap persists in the literature regarding the application of systematic and rigorous optimization methodologies designed to fine-tune these complex hybrid architectures. This gap ultimately limits their predictive accuracy when applied to demanding tasks, such as gas turbine PCD forecasting.

DOI : https://doi.org/10.25139/inform.v10i2.10226

Addressing this gap, the present study develops and validates an integrated analytical framework designed to achieve significant accuracy improvements in gas turbine PCD forecasting by employing a meticulously optimized Hybrid CNN-Residual BiLSTM model. The framework's effectiveness stems from the synergy of several core components: the inherent strengths of the hybrid architecture in processing complex sequences, domain-aware feature engineering that incorporates contextual Preventive Maintenance (PM) schedule information alongside cyclical time features, and crucially, the implementation of systematic, automated hyperparameter optimization (HPO) using the Optuna framework [19], [20]. This HPO process rigorously identifies the optimal set of hyperparameters to precisely conFig. The CNN-Residual BiLSTM model for peak predictive performance on this specific dataset. The primary contributions of this work, therefore, lie in: (1) demonstrating the tangible accuracy enhancements resulting from the application of rigorous HPO to a specific Hybrid CNN-Residual BiLSTM model within the gas turbine forecasting context; (2) integrating and evaluating the impact of domain-specific PM features on model precision; and (3) validating a comprehensive workflow that highlights how the combination of an advanced architecture, informed feature engineering, and meticulous HPO leads to superior forecasting accuracy, thereby advancing data-driven PdM capabilities. The subsequent sections detail the related work, methodology, experimental results, discussion, and conclusions of this study.

II. LITERATURE REVIEW

Forecasting key performance parameters, such as Compressor Discharge Pressure (PCD), in gas turbines presents a significant challenge due to the complex, multivariate, and non-linear nature of the underlying operational data [2]. Deep learning techniques, particularly those designed for sequential data, have emerged as powerful tools for such industrial time-series analysis. Foundational architectures, such as Long Short-Term Memory (LSTM) [8] and its bidirectional variants (Bi-LSTM) [21], are frequently employed as powerful baseline models. Their inherent capability to capture long-range temporal dependencies often allows them to outperform traditional statistical models in various predictive maintenance (PdM) applications [21][22]. However, their performance on noisy, real-world industrial data can be inconsistent, and they often require substantial, non-trivial tuning to approach optimal results, highlighting a persistent challenge in achieving high-precision forecasts [9][25].

To address the limitations of standard recurrent networks, significant research has focused on architectural refinements. One key innovation is the adaptation of residual connections, famously introduced in computer vision [14], to recurrent architectures. By creating shortcut paths for gradient flow, residual connections facilitate the training of deeper networks and mitigate the vanishing gradient problem, enabling models like Residual Bi-LSTM to learn more complex temporal patterns [15][25]. Concurrently, 1D Convolutional Neural Networks (CNNs) have proven effective at extracting localized, informative patterns and inter-sensor correlations directly from multivariate time-series data [15]. This ability to distill salient, short-term features from raw sensor readings within specific time windows is highly relevant for complex industrial systems.

Building upon these individual refinements, hybrid deep learning architectures have gained prominence, aiming to combine the strengths of different network types in a synergistic manner. Models that integrate CNN layers for initial feature extraction with recurrent layers (LSTM or Bi-LSTM) for subsequent temporal dependency modelling are becoming a state-of-the-art approach in time-series forecasting [13][16]. The underlying rationale is that CNNs can efficiently create a rich, condensed representation of local spatio-temporal features, which the recurrent layers then use to model longer-term sequential dynamics. Successful applications in analogous domains, such as energy load forecasting [16] and wind power prediction [20], have demonstrated the potential advantages of such hybrid strategies.

Beyond the core model architecture, the efficacy of any deep learning model is heavily influenced by two critical enabling methodologies: feature engineering (FE) and hyperparameter optimization (HPO). Thoughtful FE is essential for providing the model with the most informative inputs [15], which include not only standard cyclical time transformations but also the integration of domain-specific knowledge. In the context of PdM, leveraging contextual information, such as Preventive Maintenance (PM) schedules, can provide vital signals about operational cycles and performance degradation, a factor that can significantly enhance predictive accuracy [2][22]. Equally important, yet often overlooked, is a rigorous HPO process. Complex models, such as the Hybrid CNN-BiLSTM, possess numerous hyperparameters that significantly impact performance. Manual tuning is often inefficient, while advanced automated HPO methods, such as the Bayesian optimization techniques implemented in frameworks like Optuna [19], offer a more principled and efficient path to unlocking a model's full predictive potential [20].

Synthesizing these research streams reveals a clear gap. While individual advanced components—hybrid architectures, domain-aware features, and HPO tools—exist, few studies demonstrate a rigorous, holistic integration of all these elements for the specific, high-stakes task of gas turbine PCD forecasting. Much of the existing research focuses on proposing novel architectures without systematically optimizing them or enriching them with critical operational context, such as maintenance data. This study directly addresses this gap by presenting and validating an integrated framework that synergistically combines a tailored Hybrid CNN-Residual BiLSTM architecture, domain-informed PM-aware features, and meticulous, automated HPO. The objective is to demonstrate that this synergistic combination, rather than any single component alone, is key to achieving superior forecasting accuracy for data-driven gas turbine PdM.

III. RESEARCH METHODOLOGY

The methodology in this research is systematically designed to develop and rigorously evaluate a high-accuracy forecasting

DOI : https://doi.org/10.25139/inform.v10i2.10226

model for gas turbine Compressor Discharge Pressure (PCD). The process encompasses several key stages: acquisition and comprehensive pre-processing of operational data, domain-informed feature engineering, definition and systematic optimization of the core Hybrid CNN-Residual BiLSTM architecture, establishment of baseline models for comparative assessment, and the specification of quantitative evaluation metrics. The overall workflow is illustrated in Fig.1.

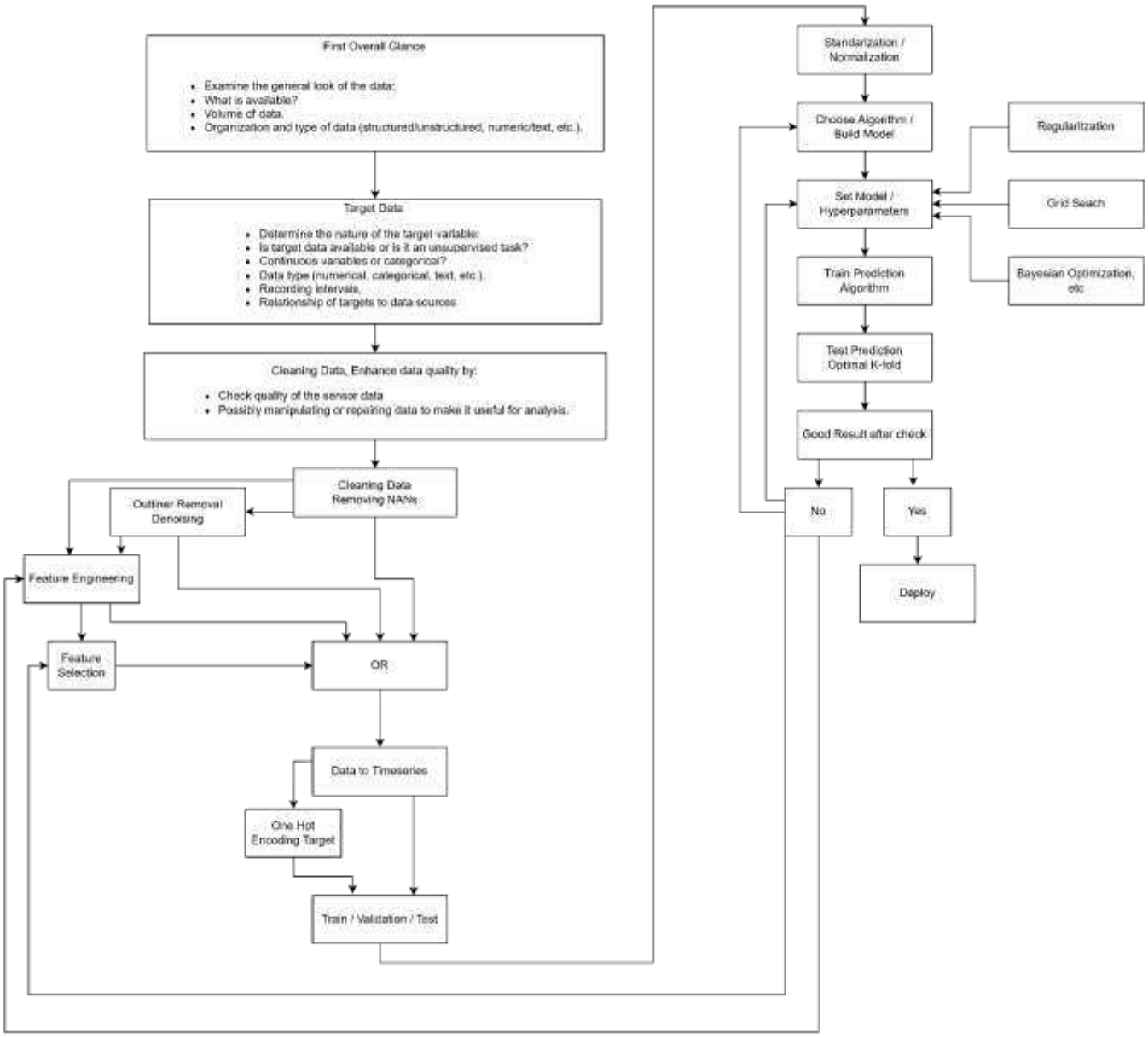

Fig.1. Systematic Methodology for Model Development and Evaluation

A. Dataset Description and Pre-processing

The empirical foundation of this work is a multivariate time-series dataset from an industrial gas turbine at a Central Processing Plant (CPP) in Eastern Indonesia. To construct a comprehensive model, this study leverages a rich, high-dimensional set of operational parameters. The raw data underwent a rigorous pre-processing pipeline to ensure its quality and suitability for modelling.

The process commenced with addressing missing values through linear interpolation. This method estimates a data point by constructing a straight line between the two nearest known data points, as governed by Equation (1), where a point $P(x,y)$ is estimated between two known points, $P_1(x_1,y_1)$ and $P_2(x_2,y_2)$.

$$y = y_1 + \frac{(x - x_1)(y_2 - y_1)}{x_2 - x_1} \quad (1)$$

DOI : https://doi.org/10.25139/inform.v10i2.10226

Subsequently, to isolate periods of stable operation, the dataset was filtered to retain records exclusively within a consistent power output range. This step was aimed at eliminating transient states such as startups and shutdowns. Duplicate entries were also identified and removed to ensure the uniqueness of each record.

Exploratory Data Analysis (EDA) was conducted to gain insights into the dataset's characteristics. The distribution of each operational variable was visualized using violin plots, as shown in Fig.2. This analysis facilitated an understanding of the central tendency, spread, and probability density of each feature. Furthermore, inter-feature correlations were examined using a heatmap, as shown in Fig.3, to identify linear relationships between variables and assess potential multicollinearity.

Finally, all selected features were normalized using Z-score standardization to rescale the data to a standard normal distribution (a mean of 0 and a standard deviation of 1). This process is governed by Equation (2), where the mean (μ) and standard deviation (σ) parameters were computed exclusively from the training data to prevent information leakage.

$$z = \frac{(x - \mu)}{\sigma} \tag{2}$$

Upon completion of the pre-processing stage, the dataset was chronologically partitioned into three subsets: a training set (70%), a validation set (20%), and a test set (10%). This division ensures that the model's generalization performance is evaluated objectively on data it has not previously encountered.

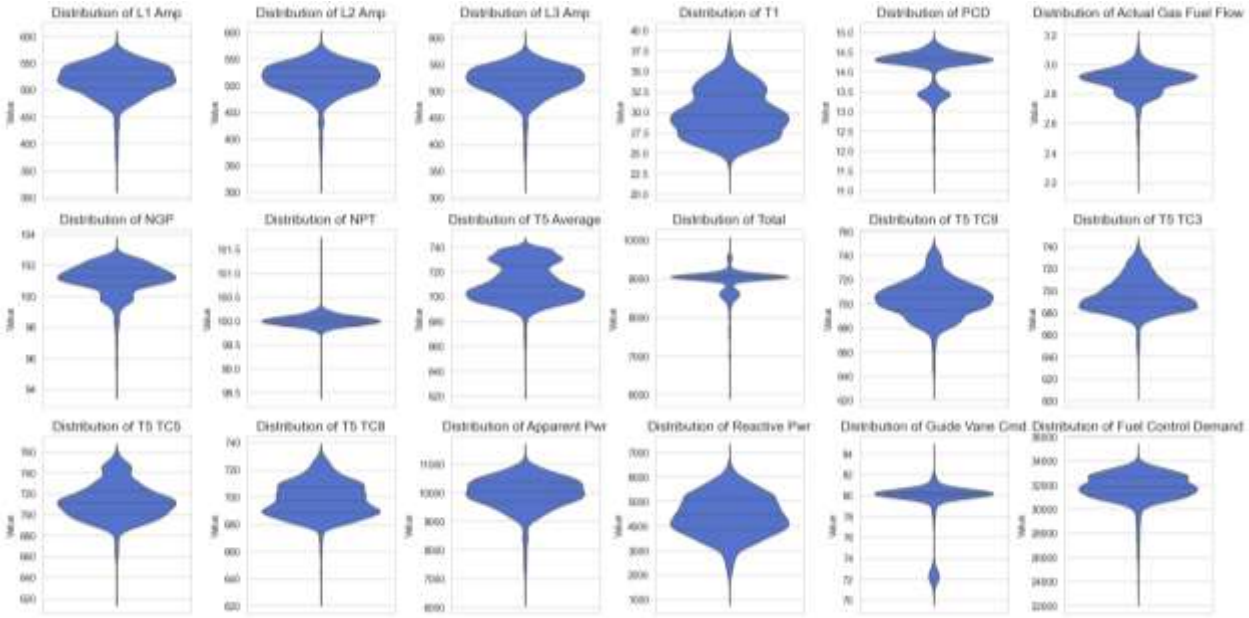

Fig.2. Distribution Analysis of Operational Features Using Violin Plots. These Plots Illustrate the Spread, Density, And Value Range for Each Key Parameter

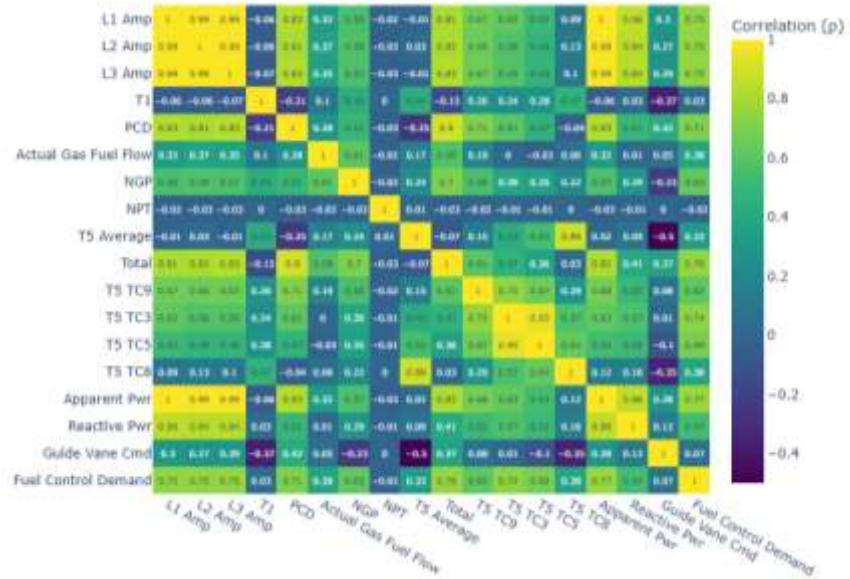

Fig. 3. Pearson Correlation Heatmap of Operational Features. Lighter Colours Indicate a Strong Positive Correlation, While Darker Colours Signify a Negative Correlation

DOI : https://doi.org/10.25139/inform.v10i2.10226

B. Feature Selection and Engineering

The methodological approach for this study is centered on a philosophy that leverages the deep learning model's intrinsic capability for representation learning. A comprehensive, high-dimensional feature space was curated, comprising 22 distinct operational parameters, to provide a multifaceted view of the gas turbine's operational state at each timestep. This strategy deliberately minimizes dependence on extensive handcrafted feature engineering, instead empowering the model to discover complex patterns from the rich data stream autonomously. The complete list of features is detailed in Table I.

While most features were sourced directly from sensors, the most impactful step in feature engineering was the creation of a contextual variable: Remaining Run Hours (Remaining_RH). This feature was specifically designed to provide the model with explicit awareness of the turbine's position within its operational and maintenance cycles. The immense value of this feature is visually demonstrated in Fig. 4, which illustrates a direct relationship between the maintenance cycle, represented by the sawtooth pattern of Remaining RH, and the behavior of critical performance parameters. As Remaining_RH increases linearly and then sharply resets to zero following a maintenance event, corresponding shifts can be observed in the Compressor Discharge Pressure (PCD) and Average Exhaust Temperature (T5 Average). For instance, after a maintenance event (indicated by a reset), the PCD signal often stabilizes at a different operational level, and the volatility of the T5 Average changes. By incorporating Remaining_RH, we equip the model with the ability to understand these cyclical performance shifts, a crucial element for achieving high-accuracy, long-term forecasting that would otherwise be lost.

To validate the relevance of the other curated sensor-based features, a quantitative analysis was performed using the Pearson correlation coefficient (visualized in the heatmap in a subsequent Fig.3). This analysis confirmed that key physical parameters, such as total power output and generator currents, exhibit a strong linear relationship with PCD, reinforcing the decision to include them in the final, comprehensive feature set.

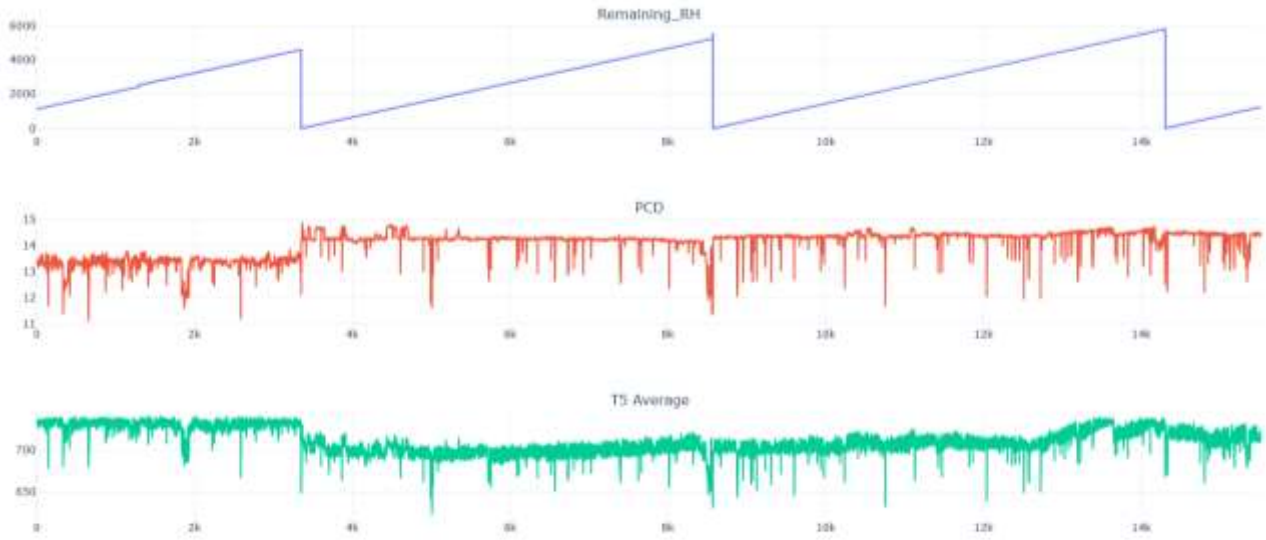

Fig.4: Visualizing the Impact of Maintenance Cycles on Key Turbine Performance Parameters

TABLE I
DETAILED BREAKDOWN OF THE COMPREHENSIVE FEATURE SET FOR TURBINE MODELLING

Feature Name	Description	Type
L1 Amp, L2 Amp, L3 Amp	Generator output current for each phase	Sensor (Electrical)
Apparent Pwr	Total power, including real and reactive	Sensor (Electrical)
Reactive Pwr	Reactive power component	Sensor (Electrical)
Total	Total Real Power Output (kW)	Sensor (Electrical)
T1	Compressor Inlet Temperature	Sensor (Thermo)
PCD	Compressor Discharge Pressure (Target)	Sensor (Thermo)
Main Vlv Disch Press	Main Valve Discharge Pressure	Sensor (Thermo)
Actual Gas Fuel Flow	Rate of fuel consumption	Sensor (Thermo)
NGP	Gas Producer Speed	Sensor (Thermo)
NPT	Power Turbine Speed	Sensor (Thermo)

DOI : https://doi.org/10.25139/inform.v10i2.10226

Feature Name	Description	Type
T5 Average	Average Exhaust Temperature	Sensor (Thermo)
T5 TC3/TC5/TC8/TC9	Individual exhaust thermocouple readings	Sensor (Thermo)
Remaining_RH	Remaining Run Hours until next maintenance	Contextual
Guide Vane Cmd	Inlet Guide Vane command signal	Sensor (Control)
Fuel Control Demand	Fuel control demand signal	Sensor (Control)
time_idx	Continuous time index from start	Time Index
row_idx	Cumulative row index	Time Index

C. Proposed Model Architecture

The core of the proposed forecasting framework is a Hybrid Convolutional Neural Network-Residual Bidirectional Long Short-Term Memory (CNN-Residual BiLSTM) model. It is specifically engineered to capture the complex dynamics within the multivariate industrial time-series data. This hybrid approach is particularly effective for forecasting complex industrial time series, as it combines the feature extraction strength of CNNs with the sequential modeling capabilities of LSTMs. The model's architecture, illustrated in Fig.5, is composed of three primary blocks designed to process an input sequence with dimensions of 22 timesteps and 22 features.

The input sequence is first processed by a Convolutional Block, which contains one or more one-dimensional convolutional (Conv1D) layers. These layers function as localized feature extractors, identifying short-term patterns and inter-feature correlations from the input data stream. An activation function and Layer Normalization follow each Conv1D layer to stabilize the learning process.

The resulting feature maps are then fed to a Recurrent Block. At the heart of this block is a Bidirectional LSTM (BiLSTM) layer, which processes the sequence from both forward and backward directions to effectively model long-range temporal dependencies and capture the full context of the sequence [8], [21]. To enhance performance and facilitate the training of a deeper network, a residual connection is integrated [14]. This mechanism enables gradients to flow more directly through the network, thereby mitigating vanishing gradient issues and allowing the model to learn more complex transformations [15][25]. Finally, a dense Output Block maps the rich temporal representations from the recurrent core to a single, continuous output value representing the final PCD forecast.

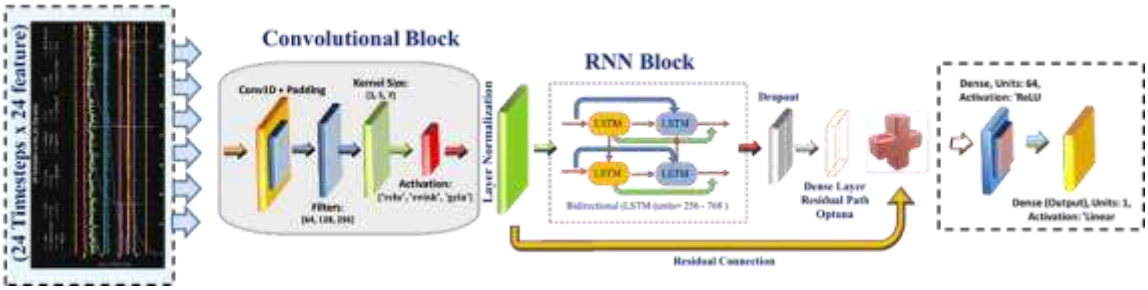

Fig.5. Proposed Hybrid CNN-Residual BiLSTM Architecture.

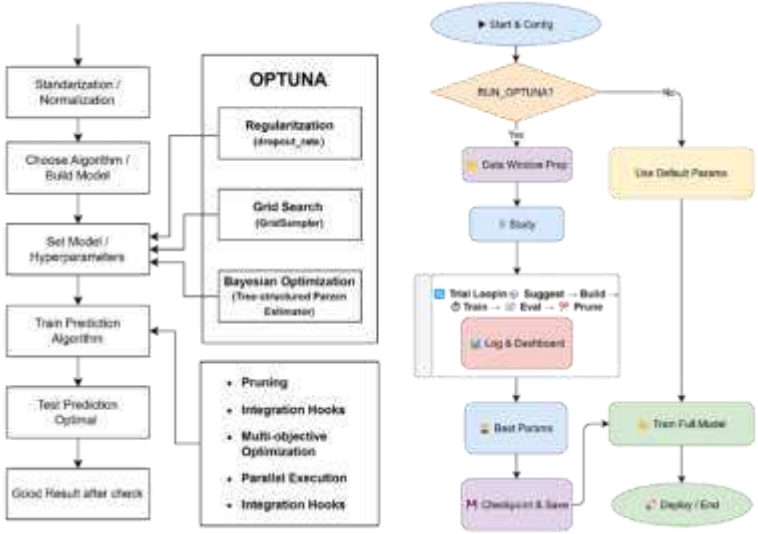

Fig.6. Optuna Hyperparameter Optimization

DOI : https://doi.org/10.25139/inform.v10i2.10226

D. Hyperparameter Optimization (Optuna)

Given that achieving optimal performance for a model of this complexity necessitates a principled and systematic approach to hyperparameter selection. Manual tuning or grid search methods are often impractical for such a high-dimensional search space. Therefore, this work employed automated hyperparameter optimization (HPO) using the Optuna framework [19], and the process is visually outlined in Fig.6.

Optuna utilizes the Tree-structured Parzen Estimator (TPE), a type of Bayesian optimization, to efficiently explore the parameter space. The core objective of the HPO process is to find the set of hyperparameters x^* from a defined search space X that minimizes a predefined objective function f(x) - in this case, the validation error (Mean Absolute Error). This optimization problem is formally expressed in Equation (3).

$$x^* = \arg\min_{x \in X} f(x) \tag{3}$$

Unlike random search, TPE constructs a probabilistic model based on past trial results, allowing it to intelligently prioritize more promising hyperparameter regions, which is highly effective for complex search spaces.

The HPO search space was carefully defined to encompass the most influential model parameters, including learning rate, network units (LSTM units and CNN filters), architectural choices (CNN kernel size and number of convolutional blocks), regularization (dropout rates), and activation functions. The optimization was conducted over a predefined number of trials, with each trial's configuration and resulting validation score being logged to a persistent SQLite database to ensure the reproducibility of the optimization process. To ensure computational efficiency, a pruning mechanism was also employed to terminate unpromising trials prematurely. This systematic process ensures that the final model configuration is objectively derived based on empirical performance rather than relying on heuristics.

E. Baseline Models

To provide a rigorous benchmark for evaluating the performance enhancements offered by the optimized Hybrid CNN-Residual BiLSTM model, several established deep learning architectures commonly used for time-series forecasting were implemented and trained under identical conditions:

1) Standard LSTM [8]: A foundational recurrent network architecture chosen to benchmark the model's fundamental ability to capture long-range temporal dependencies.

2) Bidirectional LSTM (Bi-LSTM) [21]: An enhanced architecture that processes sequences from both temporal directions, serving as a more advanced baseline for contextual understanding.

3) Residual Bi-LSTM [14][15]: A strong non-hybrid baseline that augments a Bi-LSTM with residual connections to improve training stability, representing a state-of-the-art recurrent-only model.

These baseline models were constructed with architectural complexity comparable to the proposed model, where feasible, with similar layer depths or unit counts, potentially subject to their simplified tuning. They were trained using the same pre-processed data splits (training, validation, and test) and evaluation metrics to ensure a direct and equitable comparison of performance.

F. Training

The definitive training protocol applied to the final, optimized Hybrid CNN-Residual BiLSTM model, as well as the baseline models, adhered to the following standardized setup.

1) Optimizer: The Adam optimizer [10] was employed, utilizing the specific learning rate identified as optimal by Optuna for the proposed model, as well as potentially default or separately optimized rates for the baselines. Gradient norm clipping (clipnorm=1.0) was applied to enhance training stability.

2) Loss Function: Mean Squared Error (MSE) served as the primary objective function minimized during training, as it is appropriate for the continuous nature of the PCD forecasting task. Huber loss was also considered during experimentation as a potentially more robust alternative to outliers.

3) Performance Monitoring: Progress during the training phase was carefully monitored by evaluating a series of key performance metrics. These metrics included Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2). All of these metrics were calculated on the validation dataset after each epoch concluded, in addition to continuous monitoring of the training loss value.

4) Training Duration and Regularization: The duration of the training process for each model was controlled by establishing an upper limit on the number of epochs. Early stopping was implemented as a crucial regularization technique: training was automatically halted if the monitored validation loss failed to improve for a specified number of consecutive epochs. The model weights corresponding to the epoch with the best observed validation loss were then restored and saved persistently using Model Checkpoint callbacks.

5) Batch Size: A consistent batch size of 64 samples was utilized across all training runs. This hyperparameter dictates the number of training examples propagated through the network before the model's weights are updated. Maintaining a consistent batch size is crucial for ensuring fair comparability of training dynamics and computational resource utilization across different models and experimental setups.

G. Evaluation Metrics

The predictive performance of the finally trained proposed model and all baseline models was quantitatively assessed on the unseen, held-out test set. Standard and widely accepted regression metrics were employed for this evaluation [26], [27]:

DOI : https://doi.org/10.25139/inform.v10i2.10226

1) Mean Absolute Error (MAE): Represents the average of the absolute deviations between the predicted outputs ($y^{\wedge}_i$) and the actual observed values (y_i). This indicator, calculated as shown in Equation (4), provides a straightforward and intuitive estimation of the mean magnitude of prediction error, with units consistent with the original PCD units.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y_i}\right| \quad (4)$$

2) Root Mean Squared Error (RMSE): This metric is derived by calculating the square root of the average of the squared differences between predicted values and actual observations. The squaring element in its calculation, as defined in Equation (5), causes RMSE to place a heavier penalty on large errors than MAE, making it highly sensitive to outliers.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2} \quad (5)$$

3) Coefficient of Determination (R^2): A statistical measure that reflects the percentage of variation in the dependent variable (in this case, actual PCD values) that is successfully predicted or explained by the model based on independent variable(s). This value, defined by Equation (6), indicates how well the model's predictions align with the observed data. A higher R^2 score (approaching 1) indicates that the model explains the variance in the actual data more effectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \quad (6)$$

4) Mean Absolute Percentage Error (MAPE): Presents a measure of the average absolute error, expressed as a percentage proportion of the actual values. The fundamental advantage of MAPE, as calculated according to Equation (7), is its scale-independence. However, its utility can be compromised when actual values are at or near zero [27].

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_i}}{y_i}\right| \quad (7)$$

5) Symmetric Mean Absolute Percentage Error (sMAPE): As shown in Equation (8), sMAPE is an alternative percentage error metric designed to address some limitations of MAPE, particularly asymmetry and issues with zero values. It bounds the error between 0% and 200% [27].

$$SMAPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{\left|y_i - \widehat{y_i}\right|}{\left(\left|y_i\right| + \left|\widehat{y_i}\right|\right)/2} \quad (8)$$

Crucially, all performance metrics were calculated after the model's normalized predictions were inverse-transformed back into the original PCD scale using the previously saved Z-score scaler parameters (μtrain, σtrain). This ensures that the reported error metrics directly reflect the model's accuracy in predicting the actual physical pressure values, facilitating practical interpretation. The calculation of these metrics was implemented using standard scientific libraries such as Scikit-learn and NumPy.

IV. Results And Discussion

This section presents the empirical results from implementing and evaluating the proposed forecasting framework. The experiments were designed to quantify the performance improvements achieved by the optimized Hybrid CNN-Residual BiLSTM model, with a particular focus on the impact of systematic hyperparameter optimization.

A. Experimental Setup

All experiments were conducted in a Python (v3.10) environment using TensorFlow (v2.10.1) with its Keras API for model development. Key supporting libraries included Pandas (v2.2.2), NumPy (v1.23.5), Scikit-learn (v1.5.1), and Optuna (v4.2.1). Model training and optimization were accelerated on an NVIDIA RTX 3080 TI GPU (12GB VRAM). As detailed in the methodology, the pre-processed dataset was chronologically split into training (70%), validation (20%), and test (10%) sets to ensure a robust and unbiased evaluation on unseen data.

B. Hyperparameter Optimization Results

The Optuna framework was used to optimize the architecture's key hyperparameters over 50 trials systematically. The primary objective of this process was to minimize the Mean Absolute Error (MAE) on the validation set. By leveraging the Tree-structured Parzen Estimator (TPE) sampler, the framework efficiently navigated the high-dimensional parameter space and successfully identified a superior model configuration. This optimization process was concluded by achieving a best validation MAE of 0.0378. The optimal hyperparameters determined to achieve this result are as follows: the Learning Rate is 0.003049, the LSTM units are 512, and the CNN filters are 128. CNN Kernel Size is 7, CNN Activation is 'gelu', and Dropout Rate is 0.003385.

This final set of parameters was used to conFig. The definitive model, the performance of which is evaluated in detail on the test set in Section IV-D.

C. Model Training Overview

Following the hyperparameter optimization phase, the definitive Hybrid Optuna model and all baseline architectures were trained on the combined training and validation dataset until convergence. The process was governed by the standardized protocols detailed in the methodology, including the implementation of an early stopping mechanism with a patience of 8 epochs to prevent overfitting by monitoring the validation loss.

The learning dynamics of each model throughout this training phase are visualized in Fig. 7. The plots for all evaluated metrics, including Loss, MAE, RMSE, and R^2, demonstrate stable convergence, as indicated by the flattening of both the training (solid lines) and validation (dashed lines) curves over time. A clear performance hierarchy is evident across all six plots, where the Hybrid Optuna model (red lines)

DOI : https://doi.org/10.25139/inform.v10i2.10226

consistently outperforms all baseline and non-optimized hybrid variants. Specifically, it achieves the lowest validation error values across all error-based metrics and the highest R^2 score, confirming its superior predictive accuracy.

A critical insight from Fig.7 is the model's excellent generalization capability. For the Hybrid Optuna model, the gap between the training curve (solid red line) and the validation curve (dashed red line) is minimal across all plots. This close alignment signifies that the model performs nearly as well on unseen validation data as it does on the data it was trained on, providing strong visual evidence that it has not overfit to the training set. This robust generalization underscores the effectiveness of the systematic optimization process in producing a well-calibrated and reliable forecasting model.

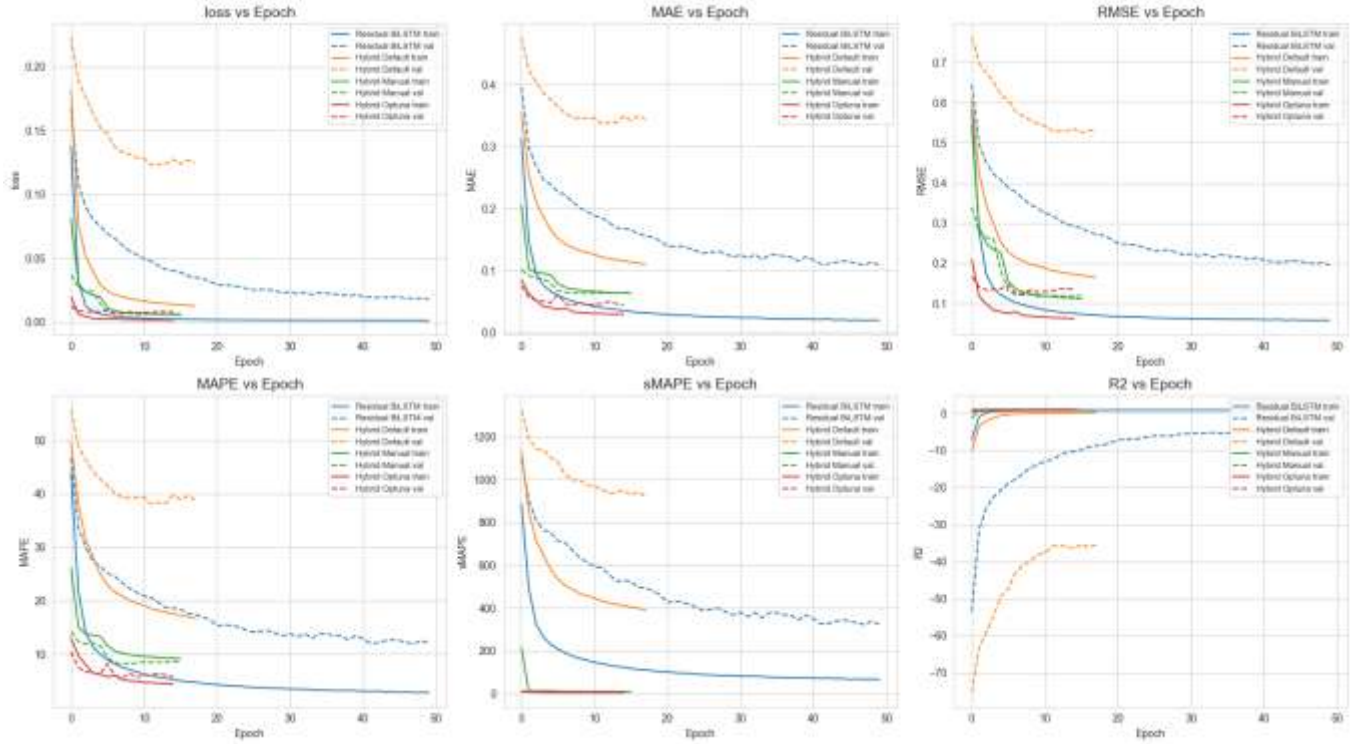

Fig.7. Training And Validation Loss Curves (For the Key Models, Demonstrating Convergence and Indicating Potential Overfitting or Underfitting).

D. Forecasting Performance Evaluation

The core evaluation of this study involved a direct comparison of the forecasting accuracy of the optimized Hybrid CNN-Residual BiLSTM model, referred to as Hybrid Optuna, against several established baseline architectures and two non-optimized variations of the hybrid model itself. The baseline models included a standard LSTM, BiLSTM, and Residual BiLSTM. At the same time, the non-optimized versions included the hybrid model with default hyperparameters (Hybrid Default) and one that was manually tuned (Hybrid Manual). The complete performance results, measured on the unseen test set, are detailed in Table II.

TABLE II
COMPARATIVE FORECASTING PERFORMANCE ON THE TEST SET

Model	Evaluation Metric				
	MAE	**RMSE**	**MAPE**	**sMAPE**	**R^2**
Linear	0.2132	0.2644	1.4905	1.5009	0.2521
LSTM	0.1351	0.2487	0.9671	0.9576	0.3387
BiLSTM	0.1495	0.1765	1.0403	1.0467	0.6670
Residual BiLSTM	0.0708	0.0850	0.4949	0.4956	0.9228
Hybrid Default Hyperparameter	0.1307	0.1538	0.9113	0.9160	0.7471
Hybrid Manual Tuned	0.0501	0.0904	0.3556	0.3559	0.9126
Hybrid Optuna	0.0298	0.0611	0.2111	0.2111	0.9601

The quantitative results unequivocally demonstrate the superior forecasting accuracy of the Hybrid Optuna model. As shown in Table II, it achieved the lowest error metrics across the board, with a Mean Absolute Error (MAE) of 0.0298 and a Root Mean Squared Error (RMSE) of 0.0611. Furthermore, it attained the highest coefficient of determination with an R^2 of 0.9601, indicating that the optimized model successfully explains 96% of the variance in the actual PCD data.

A key finding from this analysis is the significant performance gap between the optimized model and the

DOI : https://doi.org/10.25139/inform.v10i2.10226

strongest baseline, the Residual BiLSTM model. The Hybrid Optuna model improved upon the Residual BiLSTM's RMSE by approximately 28.1%, a substantial margin underscoring the impact of combining a hybrid architecture with systematic HPO. This performance differential is also visually highlighted in Fig.8, where the error bars for the Hybrid Optuna model are consistently the shortest across all metrics compared to those of every other configuration. This strong quantitative and visual evidence validates the core thesis of this work: that the synergy between an advanced architecture and meticulous optimization is essential for unlocking state-of-the-art predictive performance.

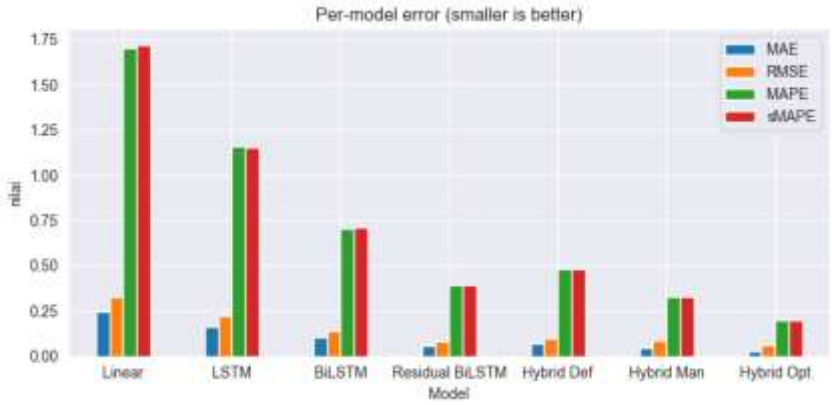

Fig.8. Comparison of Forecasting Error Metrics (MAE and RMSE) Across Models.

E. Qualitative Analysis (Prediction Visualization)

Beyond the quantitative metrics, a qualitative analysis of the model's forecasting behaviour provides crucial insights into its performance dynamics. Fig.9 presents a visual comparison between the actual PCD values (True PCD) and the forecasts generated by the key models for a selected segment of the test set. The plot clearly illustrates the superior forecasting proficiency of the Hybrid Optuna model. While other models, such as the Residual BiLSTM, struggle to capture the precise amplitude of fluctuations and exhibit a noticeable lag behind the true signal, the predictions from the Hybrid Optuna model demonstrate a remarkably close alignment with the actual PCD trend. Importantly, the model successfully captures not only the general trajectory but also the magnitude and timing of significant short-term fluctuations. This ability to accurately track the signal, even during periods of volatility, visually corroborates the strong quantitative performance reported in Table II and reinforces the robustness of the optimized framework.

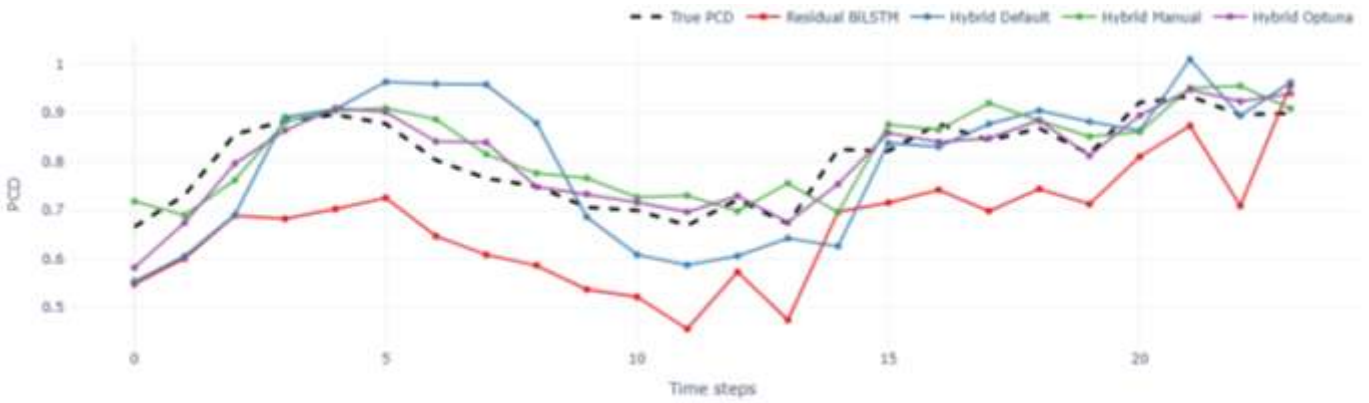

Fig.9. PCD Forecasts vs. Actual Values on Test Set Segments

V. CONCLUSION

This study successfully addressed the formidable challenge of accurately forecasting Compressor Discharge Pressure (PCD) in industrial gas turbines. We developed and validated an integrated analytical framework centered on a Hybrid CNN-Residual BiLSTM architecture. The empirical results unequivocally demonstrate that by synergistically combining this advanced architecture with comprehensive feature selection and meticulous hyperparameter optimization, the proposed model achieves a state-of-the-art forecasting accuracy (RMSE = 0.0611, R^2 = 0.9601). This performance not only significantly outperforms standard deep learning baselines but also surpasses previously reported benchmarks in similar forecasting tasks.

The principal contribution of this research lies not merely in the application of an advanced architecture, but in its holistic

DOI : https://doi.org/10.25139/inform.v10i2.10226

and rigorous methodology. Firstly, this work provides a definitive, quantifiable demonstration of the critical impact of automated hyperparameter optimization (HPO), a crucial step for unlocking peak performance that is often overlooked in applied forecasting studies. Secondly, it validates the efficacy of using a rich, multi-modal feature set, proving that providing the model with a comprehensive operational context is key to enhancing predictive precision. Finally, this research establishes a robust and reproducible end-to-end workflow that serves as a blueprint for developing high-fidelity predictive models for other critical industrial assets.

The enhanced forecasting accuracy delivered by this framework holds considerable practical implications for data-driven predictive maintenance (PdM), enabling more reliable condition monitoring and proactive maintenance scheduling. While the results are highly promising, future work should focus on validating the framework's generalizability across different turbine types and operational environments. Further exploration into explainability methods to interpret the model's decisions and model compression techniques for efficient real-time deployment is also warranted.

REFERENCES

[1] M. P. Boyce, *Gas Turbine Engineering Handbook*, 5th ed. Elsevier, 2022.
[2] M. H. Pashaei and A. Bagheri, "A comprehensive review on gas turbine performance and fault diagnosis using artificial intelligence methods," *Energy Reports*, vol. 8, pp. 5649–5678, Apr. 2022, doi: 10.1016/j.egyr.2022.04.021.
[3] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Sci Rep*, vol. 8, no. 1, p. 6085, Apr. 2018, doi: 10.1038/s41598-018-24271-9.
[4] R. J. Abinaya and G. Rajakumar, "Accurate Liver Fibrosis Detection Through Hybrid MRMR-BiLSTM-CNN Architecture with Histogram Equalization and Optimization," *Journal of Imaging Informatics in Medicine*, 2024, doi: 10.1007/s10278-024-00995-1.
[5] Z. Zare and A. Ordys, "Gas turbine prognostic and health management: A review of methods and applications," *Renewable and Sustainable Energy Reviews*, vol. 152, p. 111669, Dec. 2021, doi: 10.1016/j.rser.2021.111669.
[6] W. Zhang, D. Yang, and H. Wang, "Recent advances and trends of predictive maintenance from data-driven machine prognostics perspective," *Measurement*, vol. 176, p. 109121, Jun. 2021, doi: 10.1016/j.measurement.2021.109121.
[7] N. H. Hussin *et al.*, "Handling Volatility and Non-linearity in Wind Speed Data: A Comparative Analysis between ARIMA-GARCH and ARIMA-MLP," *Journal of Advanced Research in Applied Mechanics*, vol. 121, no. 1, pp. 44–57, Jul. 2024, doi: 10.37934/aram.121.1.4457.
[8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
[9] J. Zhang, X. Li, J. Tian, and H. Luo, "An integrated multi-head dual sparse self-attention network for remaining useful life prediction," *Reliab Eng Syst Saf*, vol. 233, p. 109096, May 2023, doi: 10.1016/j.ress.2023.109096.
[10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learning Representations (ICLR)*, San Diego, CA, USA, May 2015.
[11] Y. A. Ramaziyah and E. B. Setiawan, "Hybrid deep learning CNN and BiLSTM with FastText as feature expansion for sentiment analysis in President Election 2024," in *Proc. 2024 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, Kuta, Bali, Indonesia, Feb. 2024, pp. 176–183. doi: 10.1109/COMNETSAT63286.2024.10862946.
[12] S. Ahmad, M. Asghar, F. Alotaibi, and Y. D. Al-Otaibi, "A hybrid CNN + BiLSTM deep learning-based DSS for efficient prediction of judicial case decisions," *Expert Syst Appl*, vol. 209, p. 118318, Dec. 2022, doi: 10.1016/j.eswa.2022.118318.
[13] M. A. Dar and J. Pushparaj, "Hybrid architecture CNN-BLSTM for automatic speech recognition," in *Proc. 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)*, Coimbatore, India, Mar. 2024, pp. 1–4. doi: 10.1109/AIIoT58432.2024.10574740.
[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
[15] R. Huan, Z. Zhan, L. Ge, K. Chi, P. Chen, and R. Liang, "A hybrid CNN and BLSTM network for human complex activity recognition with multi-feature fusion," *Multimed Tools Appl*, vol. 80, no. 28–29, pp. 36159–36182, Nov. 2021, doi: 10.1007/s11042-021-11363-4.
[16] Y. Chen and Z.-P. Fu, "Multi-Step Ahead Forecasting of the Energy Consumed by the Residential and Commercial Sectors in the United States Based on a Hybrid CNN-BiLSTM Model," *Sustainability*, vol. 15, no. 3, p. 1895, Jan. 2023, doi: 10.3390/su15031895.
[17] A. Tabiti, "Neuro-Genetic System: A Hybrid System of CNN-BiLSTM Optimized by Genetic Algorithm for Road Accident Severity Prediction," in *Intelligent Systems and Pattern Recognition. ISPR 2024*, vol. 991, in Lecture Notes in Networks and Systems, vol. 991. , Springer, Singapore, 2024, pp. 29–40. doi: 10.1007/978-981-97-5495-3_3.
[18] M. T. Pawitra, H. Fakhrurroja, and L. Abdurrahman, "Predicting stock market using CNN and BiLSTM model," in *Proc. 2024 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, Bandung, Indonesia, Oct. 2024, pp. 267–272. doi: 10.1109/IC3INA64086.2024.10732245.
[19] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Anchorage, AK, USA, Aug. 2019, pp. 2623–2631. doi: 10.1145/3292500.3330701.
[20] Y. Song, Y. Wu, S. Duan, C. Dou, B. Liu, and B. Hou, "CNN-BiLSTM combined with Bayesian optimization for short-term wind power prediction," *J Phys Conf Ser*, vol. 2938, no. 1, p. 12002, Mar. 2025, doi: 10.1088/1742-6596/2938/1/012002.
[21] L. Brahimi, N. Hadroug, and A. Iratni, "Advancing predictive maintenance for gas turbines: An intelligent monitoring approach with ANFIS, LSTM, and reliability analysis," *Comput Ind Eng*, vol. 191, p. 110094, May 2024, doi: 10.1016/j.cie.2024.110094.
[22] S. M. Rezvanizaniani and T. Dempsey, "An effective predictive maintenance approach based on historical maintenance data using a probabilistic risk assessment," *Int J Progn Health Manag*, vol. 14, no. 1, p. 13, 2023.
[23] C. M. T. Yunanda, M. Hanafi, and W. M. P. Dhuhita, "Sentiment Analysis on TikTok Shop Reviews Using Long Short-Term Memory Method to Find Business Opportunity," *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, pp. 1–7, Jan. 2024.
[24] C. Yustitia and N. Charibaldi, "Chili Price Prediction One Year Ahead Using the Gated Recurrent Unit Method," *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 10, no. 1, pp. 1–7, Jan. 2025.
[25] Y. Zhang and H. Wang, "Multi-head attention-based probabilistic CNN-BiLSTM for day-ahead wind speed forecasting," *Energy*, vol. 281, p. 127865, Oct. 2023, doi: 10.1016/j.energy.2023.127865.
[26] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer, 2009.
[27] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int J Forecast*, vol. 22, no. 4, pp. 679–688, Oct. 2006, doi: 10.1016/j.ijforecast.2006.03.001

DOI : https://doi.org/10.25139/inform.v10i2.10226