

Article

Wind Tunnel Valve Opening Prediction Based on Multi-Target Mutual Information and Progressive Dilated TCN

Qiang Li ¹, Jin Guo ^{1,*}, Lin Cheng ¹, Yufei Xing ² , Daijun Ling ³, Mengqi Cong ² and Qiulin Dai ³

¹ China Aviation Industry Aerodynamics Research Institute, Shenyang 110034, China; dbsmuizi@163.com (Q.L.); chenglin_bit@163.com (L.C.)

² College of Mechanical & Electrical Engineering, Shenyang Aerospace University, Shenyang 110136, China; 20133374@sau.edu.cn (Y.X.); sacphm@163.com (M.C.)

³ China Aero-Engine Corporation Sichuan Gas Turbine Research Institute, Chengdu 610500, China; gteldj@163.com (D.L.)

* Correspondence: webbersuperman@163.com

Abstract

Accurate valve opening prediction is important for improving the stability and control performance of wind tunnel systems. To address the challenges of strong variable coupling, nonlinear dynamics, and complex temporal dependencies, this paper proposes a mutual information-based progressive dilated temporal convolutional network, named MI-PD-TCN. The proposed method first preprocesses multi-source sensor and valve operation data, constructs candidate features such as historical lag and differential features, and selects informative variables using mutual information. Then, phase space reconstruction and a sliding-window strategy are used to generate temporal input tensors. Finally, a progressive dilated TCN is developed to extract multi-scale temporal dependencies and predict valve openings. Experimental results show that the proposed method improves prediction accuracy and generalization performance, providing an effective data-driven solution for valve opening prediction in wind tunnel control systems.

Keywords: wind tunnel control system; valve opening prediction; mutual information; phase space reconstruction; temporal convolutional network

1. Introduction

Wind tunnels are critical experimental facilities for aerodynamic research, where precise regulation of flow-field parameters such as Mach number, total pressure, and temperature is essential for obtaining reliable test data [1]. In continuous wind tunnels, the flow field is regulated by coordinating multiple valves, among which the main inlet regulating valve plays a central role in adjusting the incoming airflow. The opening of this valve, together with other coupled control valves, must be precisely controlled to maintain target aerodynamic conditions under varying operating regimes and transition processes. However, the dynamic relationship between valve openings and flow-field parameters is governed by complex compressible fluid dynamics, which exhibits strong nonlinear coupling, transport delays, hysteresis, and condition-dependent behaviors. These characteristics make first-principles modeling computationally expensive and difficult to apply in real-time control [2]. Therefore, an accurate data-driven predictive model that can estimate the required valve opening directly from measurable process variables is urgently needed. Such a model can provide feed-forward compensation, reduce manual tuning efforts, and enhance the robustness of wind tunnel control systems.



Academic Editor: Bosko Rasuo

Received: 29 May 2026

Revised: 2 July 2026

Accepted: 15 July 2026

Published: 28 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Traditional industrial valve modeling approaches usually rely on mechanistic or semi-empirical formulations. Although these methods offer physical interpretability, they require extensive domain knowledge and simplifying assumptions about complex thermo-fluid processes. In wind tunnel research, Yi et al. [1] proposed a wavelet packet analysis combined with Vision Transformer for valve leakage fault diagnosis, demonstrating the potential of deep learning in wind tunnel valve systems. Zhao et al. [3] developed a bidirectional LSTM-based method for missing sensor data imputation in wind tunnel tests. However, these studies mainly address fault diagnosis and data recovery, rather than the real-time prediction of valve opening positions from operational sensor streams—a typical soft-sensing problem that remains insufficiently explored.

Over the past decade, data-driven methods have emerged as powerful alternatives for industrial process modeling. Recurrent neural networks, particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), have been widely adopted for time-series prediction [4]. Despite their success, RNN-based models suffer from inherent limitations, including non-parallelizable recurrent computations and difficulties in capturing very long-range dependencies due to gradient vanishing or exploding [5]. The Temporal Convolutional Network (TCN), introduced by Bai et al. [5], overcomes these issues by employing causal dilated convolutions, enabling efficient parallel training and an exponentially growing receptive field. TCN has shown strong performance in various industrial applications, including quality prediction and soft sensing [6–8]. However, its application to wind tunnel valve opening prediction, especially under multi-valve coupling and heterogeneous operating conditions, is still limited.

Beyond the choice of temporal backbone, industrial soft-sensing tasks face two additional critical challenges: feature selection and data imbalance. Modern wind tunnel facilities generate hundreds of sensor measurements, many of which are redundant or weakly correlated with the target valve opening. Directly using all variables increases model complexity and may degrade generalization. Mutual information (MI), an information-theoretic measure capable of detecting both linear and nonlinear dependencies, has been effectively used for feature screening in industrial processes [9,10]. However, most existing MI-based selection methods are applied independently to individual datasets or single targets. In wind tunnel systems where multiple valves are physically coupled, a feature that is informative for only one valve may be less representative of the overall operating state. Hence, a multi-target MI strategy that jointly considers several coupled valves is needed to obtain a globally relevant and robust feature set.

Moreover, wind tunnel experimental data are typically collected from multiple test campaigns with different operating conditions, resulting in severe sample imbalance among files. Training a model on such imbalanced data often leads to bias toward dominant files and poor generalization to underrepresented conditions. Sample-level weighting strategies can mitigate this, but they must be integrated with globally consistent preprocessing to enable offline training and online deployment without retraining.

To address these interconnected challenges, this paper proposes a mutual information-based progressive dilated Temporal Convolutional Network, named MI-PD-TCN, for wind tunnel valve opening prediction. The main contributions are summarized as follows.

1. A multi-target mutual information feature selection strategy is introduced. By computing MI between candidate sensor variables and three coupled valve openings (Valve 1005, Valve 1101, and Valve 1801) and averaging the scores, the method selects features that are jointly informative across multiple coupled valves. This yields a compact and physically meaningful input set that is globally relevant across heterogeneous data sources.

2. A comprehensive preprocessing pipeline is designed, integrating target power transformation for distribution calibration, global Min-Max normalization for consistent scaling across all experimental files, and phase space reconstruction with sliding-window sampling to enhance the representation of nonlinear temporal dynamics and abrupt step changes.
3. A progressive dilated TCN (PD-TCN) backbone is developed to extract multi-scale temporal dependencies. The progressive dilation structure expands the receptive field in a structured manner, allowing the model to capture both short-term fluctuations and long-range delayed responses in valve control.
4. A sample-balanced training strategy is adopted to handle severe file-length imbalance. Instance-level weights are assigned inversely proportional to the length of each source file, ensuring that short but important operating conditions contribute adequately to the optimization.
5. The proposed framework is validated on real wind tunnel operational data. Experimental results demonstrate that MI-PD-TCN achieves superior prediction accuracy and stability compared to conventional baseline models and ablation variants, confirming its practical effectiveness.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 formulates the problem and describes the wind tunnel valve system. Section 4 presents the proposed methodology in detail. Section 5 reports and discusses the experimental results. Section 6 concludes the paper and outlines future work.

2. Related Work

Building upon the challenges outlined in Section 1, this section reviews the literature along four dimensions: temporal convolutional networks for industrial prediction, information-theoretic feature selection, phase space reconstruction for nonlinear dynamics, and valve prediction combined with data imbalance handling. Finally, the research gaps are summarized, and the contributions of this paper are positioned.

2.1. Task Scope and Architectural Positioning

This work addresses the problem of predicting valve opening using short-window temporal context ($T = 10$ steps of historical sensor data). Given a short sequence of recent measurements from multiple sensors, the model predicts the current or next valve opening position—a direct regression task rather than multi-step forecasting. This problem formulation is fundamentally distinct from long-horizon forecasting ($T \gg 100$) for which recent Transformer variants (Informer, PatchTST, TimesNet) are specifically optimized.

Transformer-based models employ sparse attention and temporal decomposition to efficiently handle 100+ step-ahead predictions with complex seasonal patterns, but become computationally wasteful and over-parameterized when applied to short-window direct regression without proportional accuracy improvements. Consequently, this work focuses on CNN, RNN, and TCN-family architectures, which are well-established, practical, and computationally efficient for short-horizon soft-sensing applications. Our experimental comparison (Section 5) includes LSTM, CNN, basic TCN, and classical ML baselines (XGBoost, LightGBM), evaluated under identical conditions using the same feature set derived from mutual information filtering.

2.2. Temporal Convolutional Networks for Industrial Process Prediction

The Temporal Convolutional Network [5] leverages causal dilated convolutions to capture long-range temporal dependencies without the vanishing gradient problems of

recurrent networks. Its parallel computation capability and stable gradient flow make it attractive for industrial time-series modeling.

Several TCN variants have been proposed for quality prediction and soft sensing. Zhang et al. [6] augmented TCN with spatial and temporal attention mechanisms to adaptively weight features and time steps. Yuan et al. [7] developed a pruned tree-structured TCN that removed 91.8% of parameters while maintaining accuracy. Yao et al. [8] integrated Kolmogorov-Arnold networks with graph convolution and TCN to improve nonlinear feature extraction. Comparative studies by Lima et al. [11] and Zhou et al. [12] confirmed that TCN can match or outperform LSTM at lower training cost.

In the valve domain, Li et al. [13] proposed Trans-TCN-GRU for remaining useful life prediction of pneumatic control valves, and Wang et al. [14] applied an improved TCN with convolutional auto-encoder preprocessing for nuclear power plant valve prognostics. These studies demonstrate TCN's suitability for valve-related tasks. However, they mainly focus on prognostics or fault diagnosis rather than direct real-time opening prediction. Moreover, they do not address the coupling of multiple interacting valves or the need for globally consistent feature selection across heterogeneous wind tunnel experiments.

2.3. Information-Theoretic Feature Selection for Industrial Data

Mutual information and its normalized variant, Maximum Information Coefficient (MIC), are widely used to detect nonlinear dependencies in industrial data [9]. Unlike Pearson correlation, MI captures general statistical relationships between variables, which is critical in processes governed by complex physical mechanisms.

Pan et al. [10] employed MIC for variable screening in the Czochralski silicon crystal growth process, followed by TCN-LSTM modeling, achieving an R^2 of 0.972. Gao et al. [15] and Tuo et al. [16] combined MIC with kernel density estimation and graph-based feature aggregation within TCN frameworks for soft sensing. These works validate the effectiveness of information-theoretic feature selection prior to temporal modeling.

However, existing methods typically perform feature selection independently for each dataset or treat a single target variable in isolation. In wind tunnel control, multiple valves are coupled, and the predictive model must generalize across numerous experimental files with varying conditions. A feature set selected using only the main valve may miss variables that are jointly informative for the overall system state. Therefore, a multi-target MI strategy that averages dependency scores over several coupled valves can yield a more robust and globally consistent input representation. To the best of our knowledge, such a joint MI-based selection strategy has not been explored for wind tunnel valve prediction.

2.4. Phase Space Reconstruction for Nonlinear Dynamics

Phase space reconstruction (PSR), founded on Takens' embedding theorem [17], recovers the underlying dynamics of a nonlinear system from observed time series by constructing delayed-coordinate embeddings. Given an embedding dimension m and delay τ , PSR enriches the input with historical states, enabling models to better capture the system's evolution.

Pan et al. [18] proposed MSK-TCN, which integrates multi-scale kernels with phase space reconstruction for multivariate chaotic time series prediction. Su et al. [19] combined PSR with pipelined recurrent wavelet neural networks for effluent quality prediction. Huang et al. [20] used CNN-based feature extraction on phase-reconstructed data. These studies demonstrate that PSR can significantly improve prediction accuracy for nonlinear and non-stationary processes.

For wind tunnel valve systems, valve openings often undergo abrupt step changes during flow-condition transitions, alongside smoother drifts during steady operation.

PSR provides a principled way to embed recent historical dynamics into the input tensor, complementing hand-crafted differential features. The integration of PSR with statistical lag features to handle both gradual and abrupt dynamics has not been systematically investigated in the context of valve opening prediction.

2.5. Valve Prediction, Wind Tunnel Applications, and Data Imbalance

Valve-related prediction has attracted growing attention. Li et al. [13] and Wang et al. [14] applied TCN-based models to valve remaining useful life estimation, achieving substantial error reductions. Zheng and Lei [2] developed a TCN-FCM predictor for coordinating valve openings in district heating networks. In wind tunnel research, Yi et al. [1] proposed a leakage fault diagnosis framework for pressure-regulating valves, and Zhao et al. [3] focused on imputing missing sensor data using Bi-LSTM-NARX. These efforts confirm the utility of deep learning for wind tunnel and valve systems, but real-time prediction of control valve openings from operational sensor data remains largely unexplored.

Data imbalance across experimental campaigns is another practical concern. Wind tunnel tests generate files of very different lengths, leading to biased training if not addressed. Sample-level weighting strategies, such as inverse frequency weighting, have been used in classification [21], but their application to time-series regression with file-level imbalance in industrial soft sensing is less common. Cheng et al. [21] modeled multi-output prediction uncertainty with Gaussian mixtures, while Hssayeni and Ghoraani [22] employed conditional GANs for imbalanced time-series regression. However, these methods do not consider the joint requirement of globally consistent preprocessing, multi-target feature selection, and deployment-ready serialization that are essential for wind tunnel valve prediction.

2.6. Research Gaps and Contributions of This Paper

The literature review reveals the following gaps. (1) TCN has not been applied to real-time opening prediction of the main inlet regulating valve in wind tunnels under coupled multi-valve dynamics. (2) Multi-target MI-based feature selection across several physically coupled valves has not been used to obtain a globally consistent input space. (3) An integrated preprocessing pipeline combining target power transformation, phase space reconstruction, differential features, and sample-weighted training has not been developed for wind tunnel valve soft-sensing. (4) A systematic validation of such a framework on real wind tunnel operational data is still lacking.

To bridge these gaps, this paper presents the MI-PD-TCN framework, whose novel aspects are: (i) a three-valve averaged MI criterion for feature selection; (ii) a progressive dilated TCN that captures multi-scale temporal patterns; (iii) a sample-balanced training loss based on inverse file-length weights; and (iv) the integration of the above components into a unified, serializable pipeline suitable for offline training and online prediction.

3. Methodology

This section presents the proposed MI-PD-TCN (Mutual Information-based Progressive Dilated Temporal Convolutional Network) framework for industrial valve opening prediction. The framework is designed as an end-to-end pipeline that transforms raw, multi-source heterogeneous sensor data into accurate and physically constrained valve opening estimates. As illustrated in Figure 1, the pipeline consists of four main stages: (1) data aggregation and temporal feature construction, (2) multi-target mutual information feature selection, (3) phase space reconstruction and tensor formulation, and (4) progressive dilated TCN regression with sample-balanced training.

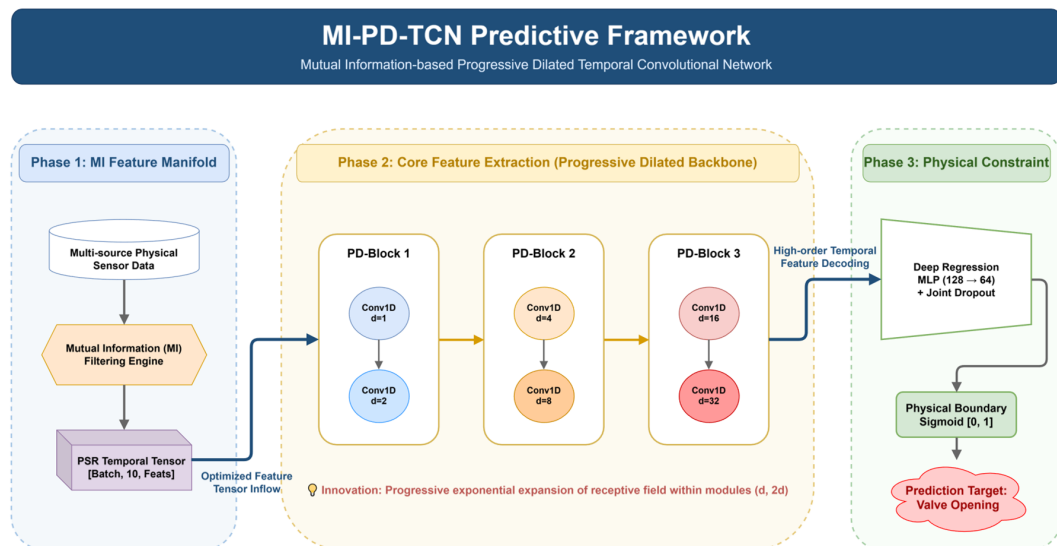


Figure 1. MI-PD-TCN Predictive Framework.

3.1. Overall Framework

Let the dataset collected from the k -th experimental file be denoted as

$$\mathcal{D}_k = \left\{ \left(\mathbf{x}_t^{(k)}, y_t^{(k)} \right) \right\}_{t=1}^{N_k} \quad (1)$$

where $\mathbf{x}_t^{(k)} \in \mathbb{R}^D$ is the D -dimensional sensor measurement vector at time step t , $y_t^{(k)}$ denotes the target valve opening, and N_k is the number of samples in the k -th file. The complete dataset is obtained by aggregating all K experimental files:

$$\mathcal{D} = \bigcup_{k=1}^K \mathcal{D}_k \quad (2)$$

In this study, the primary prediction target is the opening degree of the intake main regulating valve (Valve 1005). Two additional coupled valve sequences—Valve 1101 and Valve 1801—are jointly used during feature relevance estimation, reflecting the physical coupling among valves in the wind tunnel control circuit. The overall processing step as shown in Figure 2.

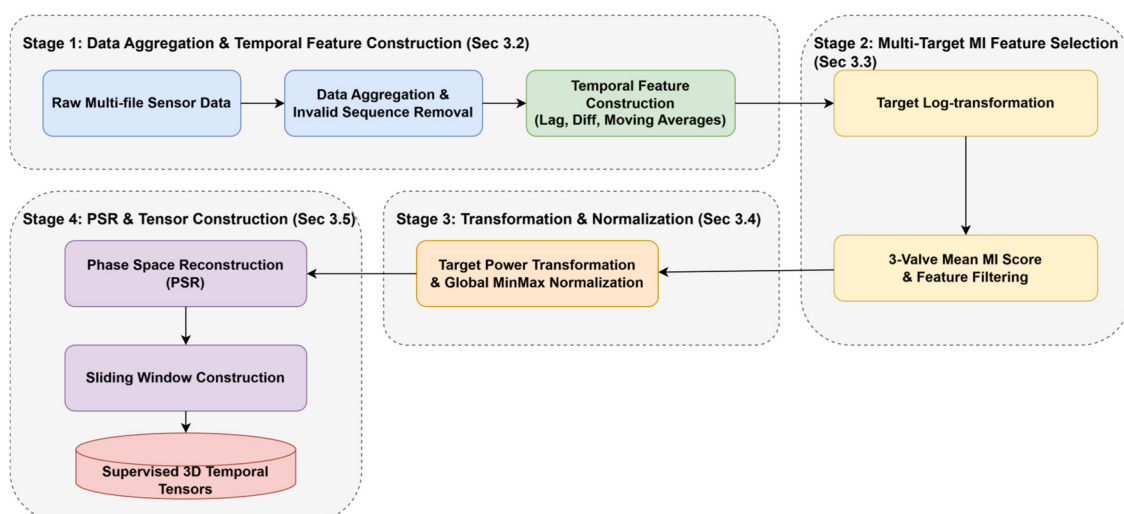


Figure 2. Data preprocessing and MI-based feature selection pipeline.

3.2. Data Aggregation and Temporal Feature Construction

The raw data consist of multiple experimental files, each corresponding to a specific wind tunnel test segment or operating condition. Before feature selection, a set of historical features is constructed from the three coupled valve opening sequences $\mathcal{Y}_v = \{y_{1005}, y_{1101}, y_{1801}\}$ to enrich the temporal representation.

For each valve opening sequence $y_v(t)$, the following five derived features are computed:

Lag feature:

$$y_{v,\text{lag1}}(t) = y_v(t-1) \quad (3)$$

First-order lagged difference:

$$y_{v,\text{diff}}(t) = y_v(t-1) - y_v(t-2) \quad (4)$$

Absolute lagged difference:

$$y_{v,\text{absdiff}}(t) = |y_v(t-1) - y_v(t-2)| \quad (5)$$

Short-term moving average:

$$y_{v,\text{ma5}}(t) = \frac{1}{5} \sum_{i=1}^5 y_v(t-i) \quad (6)$$

Long-term moving average:

$$y_{v,\text{ma20}}(t) = \frac{1}{20} \sum_{i=1}^{20} y_v(t-i) \quad (7)$$

These five features, constructed for each of the three valve sequences, provide explicit descriptions of recent valve states, instantaneous variation rates, transient magnitudes, and local steady-state trends. Missing values introduced during lag and moving-average computation are filled with zero.

Files that are too short to produce valid sliding-window samples are excluded from both training and evaluation. The minimum required sequence length is defined as:

$$N_{\min} = (m-1)\tau + T + 20 \quad (8)$$

where m is the phase space embedding dimension, τ is the delay time, and T is the temporal window length. If phase space reconstruction is not applied, the term $(m-1)\tau$ is omitted. This criterion ensures each retained file can produce sufficient temporal samples for model training and chronological data splitting.

3.3. Multi-Target Mutual Information Feature Selection

High-dimensional sensor data frequently contain irrelevant or redundant variables that increase computational burden and degrade model generalization. To address this, a mutual information (MI) based feature filtering strategy is adopted.

For two continuous random variables X_i and Y , the mutual information is defined as:

$$I(X_i; Y) = \iint p(x_i, y) \log \frac{p(x_i, y)}{p(x_i)p(y)} dx_i dy \quad (9)$$

where $p(x_i, y)$ denotes the joint probability density function of X_i and Y , and $p(x_i)$ and $p(y)$ are their respective marginal probability density functions. The ratio $\frac{p(x_i, y)}{p(x_i)p(y)}$ quantifies the degree of deviation between the joint distribution and the product of independent

distributions; a value greater than 1 indicates statistical dependence, while a value close to 1 suggests independence.

Unlike linear correlation, mutual information captures both linear and nonlinear statistical dependencies, which is particularly important for wind tunnel systems where sensor variables and valve openings may be coupled through complex aerodynamic mechanisms.

3.3.1. Logarithmic Target Pre-Transformation for MI Estimation

Before mutual information estimation, each valve opening target undergoes a logarithmic pre-transformation:

$$y_v^{\log}(t) = \log(1 + \max(y_v(t), 0)) \quad (10)$$

Industrial valve opening distributions are typically right-skewed, with values concentrated near low steady-state set points and occasional large transient excursions. The $\log(1 + y)$ mapping compresses this skewness and produces a more symmetric distribution, which enhances the robustness of the nonparametric k -nearest-neighbor MI estimator used in this study.

3.3.2. Three-Valve Mean MI Score

Rather than estimating mutual information only with respect to the primary target Valve 1005, this study adopts a multi-target mean MI strategy. For each candidate feature X_i , the MI is estimated separately against each of the three log-transformed coupled valve targets:

$$I_{i,1005} = I(X_i; y_{1005}^{\log}) \quad (11)$$

$$I_{i,1101} = I(X_i; y_{1101}^{\log}) \quad (12)$$

$$I_{i,1801} = I(X_i; y_{1801}^{\log}) \quad (13)$$

The mutual information scores are calculated exclusively from training samples to prevent data leakage. Specifically, for the final feature relevance score:

$$s_i = \frac{1}{3} \left(I^{\text{train}}(X_i; y_{1005}^{\log}) + I^{\text{train}}(X_i; y_{1101}^{\log}) + I^{\text{train}}(X_i; y_{1801}^{\log}) \right) \quad (14)$$

where superscript “train” denotes that MI estimation is computed exclusively from training samples. Feature selection threshold (15th percentile) is determined from the training distribution, and the selected feature subset is then applied consistently to validation and test sets using the same pre-fitted normalization parameters.

This multi-target aggregation reflects the physical coupling of the three valves: a feature that is jointly informative for all three coupled valves is more likely to represent the global operating state of the wind tunnel system than one that correlates with only a single valve. The selected feature subset is then defined as:

$$\mathcal{S} = \{X_i \mid s_i > \theta_{\text{MI}}\} \quad (15)$$

where θ_{MI} is the 15th percentile of the empirical MI score distribution. To reduce computational overhead, MI estimation is performed on a randomly drawn global subsample that preserves operational diversity across all experimental files.

3.4. Target Power Transformation and Global Normalization

3.4.1. Power Transformation

The distribution of valve opening values is typically highly skewed. To reduce this imbalance and improve training convergence, a power transformation is applied to each non-negative target value:

$$y' = (\max(y, 0) + \varepsilon)^\alpha \quad (16)$$

where $\varepsilon = 10^{-6}$ is a small constant for numerical stability and $\alpha = 0.4$ is the power exponent. This transformation compresses large values and expands the low-value region, yielding a more balanced target distribution.

After inference, the inverse transformation recovers the physical valve opening:

$$y = (y')^{1/\alpha} - \varepsilon \quad (17)$$

The recovered prediction is further constrained by a non-negative physical clipping:

$$y_{\text{phys}} = \max(y, 0) \quad (18)$$

This hard constraint enforces the physical requirement that valve opening cannot be negative.

3.4.2. Global MinMax Normalization

All selected input features and the power-transformed target are scaled to $[0, 1]$ using global MinMax normalization. For a selected feature x_i , the normalized value is:

$$\tilde{x}_i = \frac{x_i - x_i^{\min}}{x_i^{\max} - x_i^{\min}} \quad (19)$$

where $x_i^{\min}$ and $x_i^{\max}$ are computed from the entire globally aggregated training dataset. Global normalization is adopted instead of per-file normalization to preserve meaningful absolute physical differences across different operating conditions and experimental files.

3.5. Phase Space Reconstruction and Temporal Tensor Construction

To better represent nonlinear temporal dynamics, phase space reconstruction (PSR) is applied to the normalized input features. The micro-level PSR and sliding-window tensor construction process are illustrated in Figure 3. For a one-dimensional feature sequence $x(t)$, the phase-reconstructed state vector at time t is:

$$z(t) = [x(t), x(t - \tau), x(t - 2\tau), \dots, x(t - (m - 1)\tau)] \quad (20)$$

where $\tau = 1$ is the delay time and $m = 3$ is the embedding dimension. The reconstructed vectors of all D_s selected features are concatenated, yielding a phase-enhanced feature dimension of:

$$D_{\text{psr}} = m \cdot D_s \quad (21)$$

A sliding window of length T is then applied to generate supervised samples. For each prediction time t , the model input and label are:

$$X_t = [z(t - T), z(t - T + 1), \dots, z(t - 1)] \quad (22)$$

$$Y_t = y(t) \quad (23)$$

The final input tensor has shape:

$$\mathbf{X} \in \mathbb{R}^{B \times T \times D_{\text{psr}}} \quad (24)$$

where B is the batch size. This 3D tensor format is directly compatible with one-dimensional temporal convolution operations and enables the model to capture both short-term transitions and long-range delayed dependencies.

Phase Space Reconstruction (PSR) and Sliding-Window Sample Generation

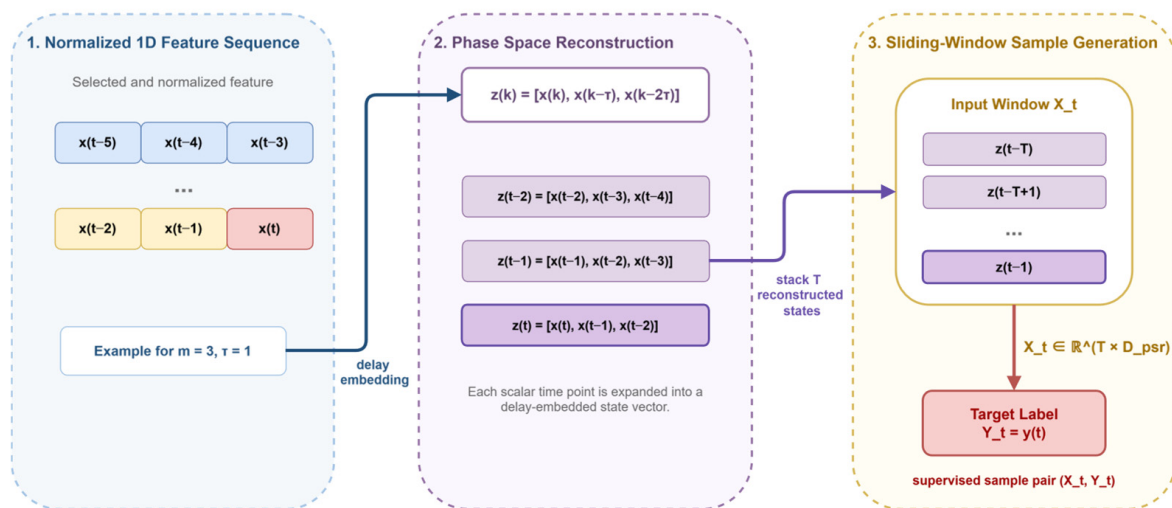


Figure 3. Phase space reconstruction (PSR) and sliding-window sample generation.

3.6. Progressive Dilated Temporal Convolutional Network

The prediction backbone is built upon Temporal Convolutional Networks (TCNs). TCNs are well-suited to time-series modeling because causal convolutions prevent future information leakage and dilated convolutions efficiently enlarge the receptive field. The proposed backbone, referred to as PD-TCN, introduces a progressive dilation design and consists of three components: an input projection layer, three progressive dilated residual blocks, and a fully connected regression head. The PD-TCN backbone is presented in Figure 4.

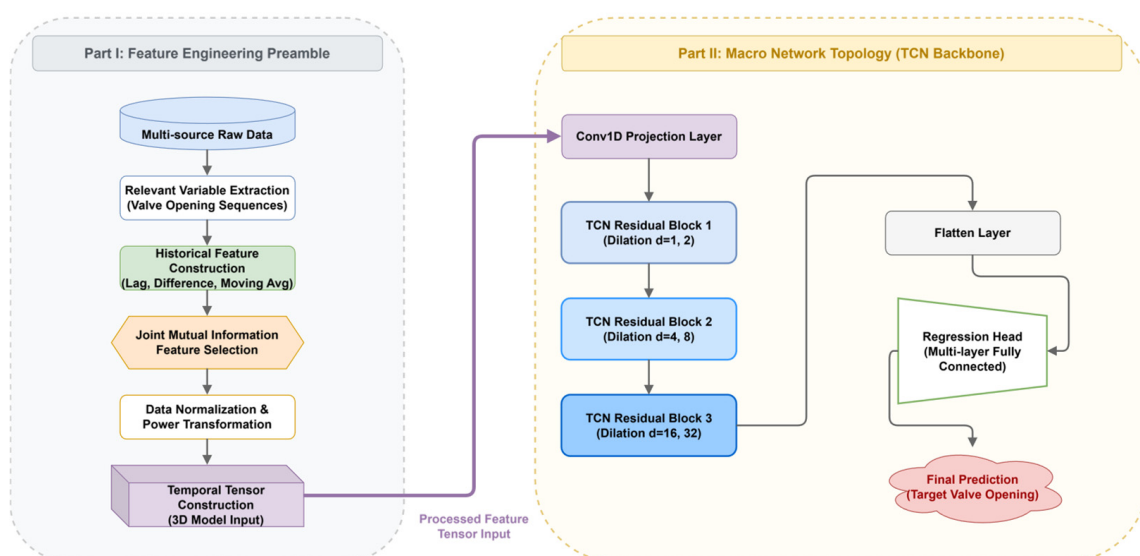


Figure 4. Architecture of the progressive dilated TCN backbone.

3.6.1. Input Projection Layer

The input tensor X is first mapped to the internal hidden channel space through a pointwise Conv1D projection:

$$H_0 = \sigma(\text{Conv1D}_{1 \times 1}(X)) \quad (25)$$

where $\sigma(\cdot)$ denotes the ReLU activation function. This layer aligns the reconstructed feature dimension D_{psr} with the uniform hidden channel width C used throughout the backbone, and performs preliminary cross-variable feature fusion.

3.6.2. Causal Dilated Convolution

For a sequence h , the dilated causal convolution at time t is defined as:

$$(h \times_d w)(t) = \sum_{i=0}^{K_c-1} w(i) h(t - d \cdot i) \quad (26)$$

where K_c is the convolution kernel size, d is the dilation factor, and w is the learnable kernel. The causal structure ensures that the output at time t depends only on current and past inputs, satisfying the requirement for real-time industrial prediction. The effective receptive field grows exponentially with dilation, enabling the model to capture delayed aerodynamic responses without a corresponding increase in parameter count.

3.6.3. Progressive Dilated Residual Blocks

The key structural contribution of PD-TCN is the progressive dilation design, which differs from conventional TCN implementations. Whereas standard TCNs assign a fixed, uniform dilation rate to both convolutions within each block and then double this rate between blocks (e.g., block 1: (1,1), block 2: (2,2), block 3: (4,4)), the proposed PD-TCN employs intra-block dilation pairs that span a wider range of scales. Specifically, the three residual blocks are assigned the following asymmetric dilation pairs, as summarized in Table 1:

Table 1. Asymmetric Dilation Pair Assignment per Residual Block.

Block	Dilation Pair (d_1, d_2)	Receptive Field Focus
Block 1	(1, 2)	Local transitions
Block 2	(4, 8)	Medium-range dependencies
Block 3	(16, 32)	Long-range delayed effects

This intra-block asymmetry allows each block to simultaneously span a wider range of temporal scales before the receptive field is expanded further at the next block level. Consequently, the receptive field grows faster and more densely across nested timescales—rather than uniformly doubling—creating a more nuanced multi-scale temporal hierarchy.

Physical Rationale: Wind tunnel valve opening dynamics exhibit two coexisting temporal behaviors: (1) fast transient step changes that follow control commands or sensor disturbances (occurring within 1–4 time steps), and (2) slower, delayed steady-state drift caused by upstream pipeline and plenum dynamics (spanning 10–50+ time steps). By pairing a “fine” intra-block dilation (e.g., $1 \rightarrow 2$ in Block 1) with a “coarse” one (e.g., $4 \rightarrow 8$ in Block 2), each block captures information at both rapid and sluggish timescales before proceeding to the next level. This hierarchical multi-scale design is well-matched to the heterogeneous timescale structure revealed by mutual information analysis (Section 4.4), which shows that both lag-1 features and 5–20 step moving averages are informative.

Within the l -th block, the feature transformation is:

$$\mathbf{U}_l = \text{Conv1D}_{d_{l,2}} \left(\sigma \left(\text{Conv1D}_{d_{l,1}} (\mathbf{H}_{l-1}) \right) \right) \quad (27)$$

Layer normalization is then applied to stabilize the hidden feature distribution across varying operating conditions:

$$\mathbf{V}_l = \text{LayerNorm}(\mathbf{U}_l) \quad (28)$$

A residual connection is added to facilitate gradient propagation through deep stacked blocks:

$$\mathbf{H}_l = \mathbf{H}_{l-1} + \mathbf{V}_l \quad (29)$$

Note: Dropout is not applied inside the residual blocks. Regularization via dropout is applied exclusively in the fully connected regression head described below.

Ablation Validation: The contribution of the progressive dilation structure is validated in Section 5.5.5 (Experiment E5), where replacing the intra-block asymmetric scheme with conventional uniform dilation $(1,1) \rightarrow (2,2) \rightarrow (4,4)$ causes average RMSE to increase from 3.1932 to 4.0583 and average R^2 to decrease from 0.9537 to 0.9265, confirming that the progressive design yields meaningful accuracy improvements for this application.

3.6.4. Regression Head

After the final progressive dilated block, the temporal feature map is flattened:

$$\mathbf{q} = \text{Flatten}(\mathbf{H}_L) \quad (30)$$

The flattened representation passes through two fully connected layers with interleaved dropout:

$$\begin{aligned} \mathbf{r}_1 &= \sigma(\mathbf{W}_1 \mathbf{q} + \mathbf{b}_1), & \mathbf{r}'_1 &= \text{Dropout}_{p=0.2}(\mathbf{r}_1) \\ \mathbf{r}_2 &= \sigma(\mathbf{W}_2 \mathbf{r}'_1 + \mathbf{b}_2), & \mathbf{r}'_2 &= \text{Dropout}_{p=0.1}(\mathbf{r}_2) \end{aligned} \quad (31)$$

The normalized output prediction is obtained as:

$$\hat{y}_{\text{norm}} = g(\mathbf{W}_o \mathbf{r}'_2 + \mathbf{b}_o) \quad (32)$$

In the complete MI-PD-TCN model, $g(\cdot)$ is the sigmoid function:

$$g(a) = \frac{1}{1 + e^{-a}} \quad (33)$$

This constrains $\hat{y}_{\text{norm}} \in (0, 1)$, consistent with the MinMax-normalized target range. In the ablation experiment E7, $g(\cdot)$ is replaced by a linear identity function to evaluate the effect of this output range constraint.

3.7. Sample-Balanced Training Strategy

Experimental files vary substantially in length. Direct uniform training would cause the model to be dominated by long files and underfit to shorter but operationally distinct test conditions. To address this, a sample-balanced weighting strategy is adopted.

Let N_k be the number of samples in the k -th file and $N_{\max} = \max_k N_k$. The weight assigned to samples from file k is:

$$w_k = \min \left(\frac{N_{\max}}{N_k}, w_{\max} \right) \quad (34)$$

where $w_{\max} = 5$ is an upper bound preventing excessively large weights for very short files. The weighted training loss is:

$$\mathcal{L} = \frac{1}{\sum_{i=1}^K N_k} \sum_{k=1}^K \sum_{i=1}^{N_k} w_k \uparrow(y_i^{(k)}, \hat{y}_i^{(k)}) \quad (35)$$

where $\uparrow(\cdot)$ is the Huber loss. The Huber loss is selected for its robustness to outliers relative to MSE while maintaining smooth gradients near the optimum:

$$\uparrow_{\delta}(\hat{y}, y) = \begin{cases} \frac{1}{2}(\hat{y} - y)^2 & \text{if } |\hat{y} - y| \leq \delta \\ \delta \left(|\hat{y} - y| - \frac{\delta^2}{2} \right) & \text{otherwise} \end{cases} \quad (36)$$

The model is optimized using the Adam optimizer with a learning rate of 5×10^{-4} . Gradient clipping with a global norm threshold of $\|g\|_2 \leq 1.0$ is applied to stabilize training of the residual dilated convolutional network. Early stopping and learning rate reduction on validation loss plateau are also employed.

4. Experiment and System

4.1. Experimental System and Data Acquisition

The experimental data used in this study were collected from a high-performance wind tunnel testbed facility. The system is a complex integrated control system consisting of the following subsystems: (1) wind tunnel body with test section and exhaust; (2) multi-stage air supply pipeline with pressure and temperature regulation stages; (3) exhaust and secondary flow system with vacuum extraction capability; (4) multiple electronically controlled regulating valves for flow and pressure modulation; (5) a distributed sensor network for real-time process monitoring; (6) a programmable logic controller (PLC) for supervisory control; and (7) an industrial computer for data logging and safety monitoring.

The structure of the wind tunnel control and data acquisition system is shown in Figure 5.

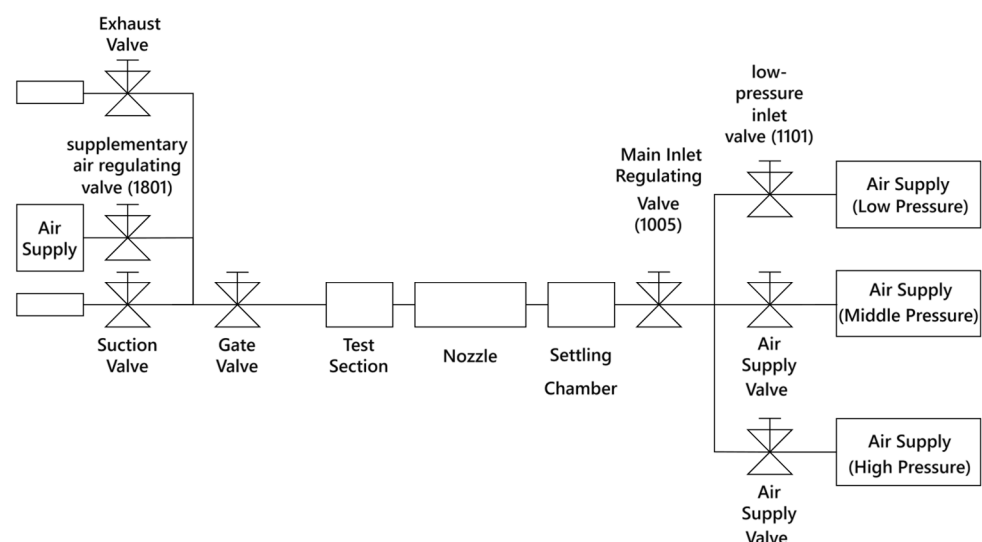


Figure 5. Experimental wind tunnel control and data acquisition system.

4.1.1. Measured Process Variables

The industrial computer continuously records 108 distinct process variables at a fixed sampling interval. Table 2 systematically lists the key process variables by functional category.

Table 2. Main categories of measured process variables.

Category	Count	Representative Variables	Physical Meaning
Air source pressure	4	P1, P2, P4	Multi-stage supply pressures
Wind tunnel pressure	3	P5, P8, Nozzle exit	Settling chamber, test chamber, nozzle
Suction system pressure	4	P11, P15, P16, P17	Secondary suction and exhaust pressures
Secondary flow pressure	4	P6, P7, P18, P19	Low-pressure circuit supply and branches
Temperature	9	T3, T4, T5, T9, T11, T15, T16, T17, wall temps	Air source, settling, exhaust, suction circuits
Target valve openings	3	Valve 1005, 1101, 1801	Main inlet (PREDICTION TARGET), low-pressure inlet, supplementary air
Auxiliary valve commands	20	Exhaust, inlet, shutoff, nozzle control valves	Flow routing and pressure regulation
Control and actuator signals	16	PID gains (P, I, D), mode flags, nozzle position, Mach number	Active control parameters and state flags
Hydraulic system	6	Pump temperature (3 stations), cooling pressure	Actuator drive fluid health
Lag-derived features	15	Position lag, velocity, moving averages (3 valves $\times$ 5 per valve)	Temporal history of target valves
Total	84	—	—

4.1.2. Physical System and Control Context

The three valve targets (1005, 1101, 1801) are the primary control actuators that regulate airflow through different circuits of the wind tunnel:

- Valve 1005 (Main inlet regulating valve): Controls primary pressurized air flow to the main supply pipeline; directly couples to settling chamber pressure (P5) and test chamber pressure (P8).
- Valve 1101 (Low-pressure inlet valve): Modulates secondary low-pressure flow; couples with low-pressure plenum dynamics.
- Valve 1801 (Supplementary air regulating valve): Regulates additional make-up air-flow; influences system-wide mass balance and transient response.

Each valve's opening must be predicted in real time to support closed-loop pressure and flow control algorithms. The prediction task is challenging due to (1) complex coupling: opening valve 1005 simultaneously perturbs pressures P5, P8, P9, and temperatures T5, T9 via thermodynamic feedback; (2) nonlinear dynamics: the flow-pressure relationship is governed by choked-flow conditions and valve flow characteristics that are highly nonlinear; and (3) multi-scale temporal behavior: valve commands produce both fast transient steps and slow relaxation governed by plenum filling dynamics (time constants ranging from <1 s to ~ 10 s).

Although three regulating valves operate in the system—Valve 1005 (main inlet), Valve 1101 (low-pressure inlet), and Valve 1801 (supplementary air)—they are tightly coupled through shared air chambers and pressure regulators. This work focuses on predicting the opening degree of Valve 1005 as the primary prediction target, as this valve directly controls the wind tunnel's main incoming flow, exhibits the most complex dynamic response, and requires the highest control precision.

For feature selection, a multi-target mutual information strategy is adopted that jointly considers all three coupled valves. Rather than using MI scores computed only with respect to Valve 1005, the method averages MI values across all three valves. This “wide-intake” design reflects a fundamental principle: a feature jointly informative for all three coupled valves is more likely to represent the global aerodynamic state of the wind tunnel system

than one correlating with only a single valve. The resulting feature set is therefore globally relevant and robust across heterogeneous experimental conditions, while the prediction model remains tightly focused on the high-precision single target (Valve 1005).

4.2. Dataset Description

The dataset consists of multiple Excel files collected from the wind tunnel experimental platform. Each file corresponds to a continuous operation sequence under a specific experimental condition. To preserve the temporal continuity of the industrial process, each file was processed independently during sample construction, and the samples were split chronologically into training, validation, and testing subsets.

The primary prediction target is the opening of Valve 1005, denoted as y_{1005} . Valve 1101 and Valve 1801 were not used as final model outputs, but were included in the feature selection stage. This design allows the selected input variables to reflect the common operating relevance of multiple coupled regulating valves rather than being determined only by a single target valve.

Before feature selection, lag-related variables were constructed for the three valve opening signals. For each valve opening variable, the following historical features were generated:

$$y_{t-1}, \Delta y_{t-1}, |\Delta y_{t-1}|, MA_5(y_{t-1}), MA_{20}(y_{t-1})$$

where y_{t-1} is the one-step lagged valve opening, Δy_{t-1} is the first-order difference in the lagged signal, $|\Delta y_{t-1}|$ represents the absolute variation, and MA_5 and MA_{20} denote the moving averages with window sizes of 5 and 20, respectively. These lag-based variables provide additional dynamic information about recent valve movements.

After removing the time column, original valve output columns, unnamed columns, and non-numeric variables, the remaining numerical variables were used as candidate input features.

4.3. Data Preprocessing

Per-file temporal splitting: Each experimental file was first split chronologically into training (50%), validation (20%), and testing (30%) subsets.

Training-only normalization: Min-max normalization parameters (min, max values) were computed exclusively from the combined training samples across all files. These fitted parameters were then applied to validation and test data without re-fitting.

Training-only feature selection: Mutual information scores were calculated only from the training subset. Features ranked above the 15th percentile in training MI were selected, ensuring feature relevance reflects only training-set statistics.

This strict separation ensures that validation and test set statistics do not influence model development, preserving the integrity of performance evaluation

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (37)$$

where x is the original feature value, and $x_{\min}$ and $x_{\max}$ are the minimum and maximum values calculated from the experimental data.

For the target valve opening, a power transformation was applied before normalization:

$$y' = (y + \epsilon)^\alpha \quad (38)$$

where y is the original Valve 1005 opening, $\epsilon = 10^{-6}$ is a small constant used to avoid numerical instability, and α is the power exponent. This transformation compresses large target values and helps stabilize model training when the target distribution is uneven.

After prediction, the inverse transformation was applied to recover the predicted value in the original valve-opening scale:

$$\hat{y} = (\hat{y}')^{1/\alpha} - \epsilon \quad (39)$$

where $\hat{y}'$ is the predicted value in the transformed space and $\hat{y}$ is the final predicted valve opening.

Phase space reconstruction was then used to enhance the temporal representation of the input variables. For a univariate time series $x(t)$, the reconstructed state vector is defined as:

$$X_t = [x(t), x(t - \tau), x(t - 2\tau), \dots, x(t - (m - 1)\tau)] \quad (40)$$

where τ is the delay time and m is the embedding dimension. In this study, $\tau = 1$ and $m = 3$. After phase space reconstruction, a sliding window with length $T = 10$ was used to construct the temporal input samples for the prediction model.

For each experimental file, the samples were divided chronologically. The first 50% of samples were used for training, the following 20% were used for validation, and the final 30% were used for testing. No random shuffling was applied before splitting, in order to avoid temporal information leakage.

4.4. Feature Selection Results and Physical Interpretation

4.4.1. Feature Screening Outcomes

The mutual information-based feature screening, applied to 103 candidate variables using the three-valve mean MI score (Equation (14)), resulted in the selection of 87 features (84.5% retention rate). The 16 filtered features had MI scores below the 15th percentile threshold and were classified as low-relevance signals. These excluded variables primarily consisted of redundant secondary sensors and rarely actuated safety valves, which are justified to remove as they contribute minimal information while increasing input dimensionality and potential noise. The detailed classification, quantity, retention rate, and physical interpretation of the retained 87 screened features grouped by functional categories are summarized in Table 3.

Table 3. Presents the composition of the 87 selected features categorized by functional role.

Functional Category	Count	Retention Rate	Physical Significance
Air source pressures (P1, P2, P4)	3	100%	Direct control input to wind tunnel
Wind tunnel key pressures (P5, P8, nozzle static, suction inlet/outlet)	7	100%	Aerodynamic state of test chamber
Secondary flow pressures (supply, branch lines)	4	100%	Cross-coupled dynamic effects
Temperatures (air source, settling chamber, wall, exhaust, suction)	9	100%	Thermodynamic effects on air density and valve response
Target valve current openings (1005, 1101, 1801)	3	100%	Direct feedback for prediction
Auxiliary control valves (22 regulators, shutoff valves)	20	91%	Indirect influence through pressure network
PID parameters and control mode signals (gains, cut-in flags, Mach mode)	12	86%	Closed-loop control characteristics
Hydraulic system states (pump outlet temp, cooling water pressure)	6	100%	Actuator response quality and hysteresis
Temporal lag features (5 features $\times$ 3 valves: lag-1, diff, abs-diff, MA5, MA20)	15	100%	Historical state and rate-of-change information

4.4.2. Physical Justification for Complete Feature Retention

Pressure subsystem (14 variables, 100% retention):

All measured pressures—from air source to test chamber to exhaust—are retained because they form a single shared gas network. Each pressure point reflects the state of the entire system due to compressibility and transient wave propagation. Removing any single pressure sensor would break the aerodynamic coupling, making it impossible for the model to infer the global flow state.

Temperature subsystem (9 variables, 100% retention):

Although multiple temperature sensors are deployed, each samples a different region (source, settling chamber, test article, walls) that influences air density locally. The compressible flow equations show that temperature variations affect aerodynamic forces and actuator response characteristics. Full retention ensures the model can capture thermodynamic feedback loops.

Three target valve openings (100% retention):

These are the primary feedback signals from the control system. The inclusion of all three valves' current positions, rather than only Valve 1005, reflects their mutual coupling through shared control networks and pressure regulators. This cross-valve state information is essential for the model to understand the coupled dynamics.

Auxiliary control valves (20 of 22, 91% retention):

The two filtered auxiliary valves are rarely actuated safety shutoff valves with nearly constant positions (open or closed). Their fixed state contributes negligible information. The remaining 20 auxiliary valves represent the full path of pressure and flow regulation affecting Valve 1005's operating point.

PID parameters and mode signals (12 of 14, 86% retention):

The selected parameters capture the closed-loop control strategy (proportional, integral, derivative gains) and mode transitions (normal operation, special Mach number control). Two rarely triggered emergency signals were filtered. These variables are critical because Valve 1005 is not an open-loop actuator—its dynamics are shaped by the controller's parameterization.

Hydraulic system states (6 variables, 100% retention):

All hydraulic subsystem measurements are retained because the control valve actuators are hydraulically powered. Variations in pump outlet temperature and cooling water pressure affect hydraulic fluid viscosity, which in turn modulates the valve's response speed and hysteresis characteristics.

Temporal lag features (15 variables, 100% retention):

The five derived features per valve (lag-1, first-order difference, absolute difference, MA-5, MA-20) encode both historical state and rate-of-change information. Their 100% retention rate validates the design choice of phase space reconstruction and the temporal convolutional architecture: the model requires explicit access to recent history to capture the nonlinear dynamics of wind tunnel valve control.

4.4.3. Implications for Model Architecture

The 100% retention of temporal lag features is particularly noteworthy. This result suggests that:

1. Temporal structure matters: The model cannot rely solely on current sensor values; it must access recent history to make accurate predictions.
2. Moving-average smoothing is informative: The inclusion of 5-step and 20-step moving averages indicates that wind tunnel valve dynamics span multiple timescales—from short-term fluctuations to long-range trends.

3. PSR design is justified: Phase space reconstruction with embedding dimension $m = 3$ and delay $\tau = 1$ (which effectively reconstructs the last 3 consecutive time steps) aligns with the natural timescale structure revealed by MI analysis.

These findings provide post hoc validation that the progressive dilated TCN architecture—which uses receptive fields spanning 1, 2, 4, 8, 16, and 32 time steps—is well-matched to the underlying temporal dependencies discovered through mutual information analysis.

4.5. Experimental Implementation Procedure

Since the overall data processing and model pipeline have been illustrated in Section 3, this section describes the experimental implementation procedure in text form.

The implementation process is summarized as follows.

Step 1—Data Loading and Chronological Splitting

All experimental Excel files were loaded from the data folder. For each file, the required valve opening columns (Valves 1005, 1101, and 1801) were verified. Files missing required variables or containing insufficient samples were excluded. For each valid experimental file, samples were chronologically divided before any preprocessing:

- Training subset: samples 1 to 50%
- Validation subset: samples 50% to 70%
- Testing subset: samples 70% to 100%

Step 2—Lag Feature Construction (Training-Only)

Lag-based features were constructed for the three valve openings to capture historical dynamics. For each training subset, the following lag features were generated:

- One-step lagged values
- Lagged first-order differences (rate of change)
- Absolute lagged differences
- Moving averages with window sizes of 5 and 20 samples

Crucially, lag features were constructed only within the training subset of each file. Lag computations did not cross the boundaries between training, validation, and testing subsets to avoid temporal information leakage. After lag feature construction, the training subsets from all files were concatenated to form a combined training dataset for the next preprocessing stages.

Step 3—Feature Selection (Training-Only)

Candidate numerical features were extracted by removing:

- Time columns
- Original (target) valve opening columns
- Unnamed or non-numeric variables

Mutual information between each candidate feature and all three valve openings was calculated exclusively using the combined training dataset. Features whose multi-target MI score exceeded the 15th percentile threshold were retained as the final input feature set. This ensures that feature selection reflects only training-data relationships, not validation or test-data statistics.

Step 4—Normalization and Target Transformation (Training-Only)

Normalization parameters were fitted using only the training data:

- Input features: Min-max scaler parameters (minimum and maximum values) were computed from the combined training dataset
- Target variable (Valve 1005): A power transformation was applied to normalize the distribution, followed by min-max scaling to the range $[0, 1]$, with parameters fitted from training data only

Step 5—Phase Space Reconstruction and Sample Construction

The normalized training, validation, and testing data were independently processed:

1. For each subset, phase space reconstruction was applied using an embedding dimension $M = 3$ and time delay $\tau = 2$, generating multi-dimensional temporal vectors from each univariate feature
2. Sliding window sampling with window size 15 was then applied to create temporal input tensors
3. Training samples from all files were combined into a single training batch for model optimization
4. Validation and testing samples were kept separate and used exclusively for their respective purposes

Step 6—Model Training and Evaluation

The MI-PD-TCN model was trained on the combined training dataset using the Adam optimizer with a learning rate of 1.995×10^{-5} , batch size of 64, and early stopping based on validation loss (patience: 12 epochs). The validation set monitored model generalization during training and triggered early termination to prevent overfitting.

After training convergence, model predictions on the test set were inverse-transformed back to the original valve-opening scale using the pre-fitted scalers. Final performance was evaluated using four metrics:

- MAE: Mean absolute error
- RMSE: Root mean squared error
- R^2 : Coefficient of determination

4.6. Evaluation Metrics

Four regression metrics were used to evaluate the prediction performance: mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination R^2 .

The MAE is defined as:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (41)$$

The RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (42)$$

The coefficient of determination R^2 is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (43)$$

where y_i is the measured valve opening, $\hat{y}_i$ is the predicted valve opening, $\bar{y}$ is the mean of the measured valve opening, and N is the number of testing samples.

MAE and RMSE measure the absolute prediction error in the original physical scale. RMSE penalizes large errors more heavily than MAE due to the squared term. R^2 evaluates the overall fitting ability of the model. A higher R^2 value closer to 1 indicates better model performance.

4.7. Hyperparameter Optimization

To ensure systematic and reproducible justification of critical hyperparameters, we conducted Bayesian hyperparameter optimization using Optuna's Tree-structured Parzen Estimator (TPE) algorithm. Over 30 trials, we optimized five core parameters: sliding window length T , phase space reconstruction delay τ , embedding dimension m , power

transform exponent p , and learning rate η . The search spaces were given in Table 4. After searching the full hyperparameter space and identifying the optimal combination, we compared the prediction performance of the optimized model against the original empirical baseline, and the quantitative results of model evaluation metrics including R^2 , MAE and RMSE are listed in Table 5, which clearly reveals significant performance gains brought by hyperparameter tuning.

Table 4. Hyperparameter Search Space.

Hyperparameter	Search Range
T (Sequence Length)	5, 10, 15, 20, 25
τ (Time Delay)	1, 2, 3, 4
m (Embedding Dimension)	2, 3, 4, 5
p (Dropout Ratio)	0.2 to 0.8
η (Learning Rate)	10^{-5} to 10^{-3}

Table 5. Performance Comparison.

Metric	Empirical Baseline	Optimized	Improvement
R^2	0.9190	0.9615	+0.0425
MAE	2.7506	1.9078	−30.6%
RMSE	4.2573	2.9507	−30.7%

This 4.6% improvement in R^2 and 30.6% reduction in MAE validate the necessity of systematic hyperparameter optimization and demonstrate that empirical selection leaves room for improvement.

4.8. Parameter Settings

The experiments were conducted on a desktop workstation running Windows 10 (Microsoft Corporation, Redmond, WA, USA) with an Intel Core i5-9400F processor (Intel Corporation, Santa Clara, CA, USA), 16 GB RAM, and an NVIDIA GeForce GTX 1660 GPU with 6 GB VRAM (NVIDIA Corporation, Santa Clara, CA, USA). The implementation was developed in PyCharm 2026.1.4 (JetBrains s.r.o., Prague, Czech Republic) and executed in Python 3.9 (Python Software Foundation, <https://www.python.org>, accessed 17 July 2026). NumPy v1.24.3 and Pandas v2.0.3 were used for data processing, scikit-learn v1.3.0 for mutual information calculation and normalization, and TensorFlow 2.13.0 (with Keras API) for building and training the neural network model. Table 6 summarizes the complete experimental parameter settings, listing the original empirical values, optimized candidates obtained from hyperparameter searching, and the final adopted configuration used for all subsequent model training and testing.

During training, early stopping was used to prevent overfitting. If the validation loss did not improve for 12 consecutive epochs, the training was terminated and the best model weights were restored. In addition, the learning rate was reduced by a factor of 0.5 when the validation loss did not improve for 6 consecutive epochs, with a minimum learning rate of 1×10^{-6} .

To reduce the influence of file-length imbalance among different experimental sequences, sample weighting was applied during model training. The weight of samples from the j -th file was calculated as:

$$w_j = \min\left(\frac{L_{\max}}{L_j}, 5.0\right) \quad (44)$$

where L_j is the number of samples in the j -th experimental file and $L_{\max}$ is the maximum file length among all files. This strategy increases the training contribution of shorter experimental sequences and prevents the optimization from being dominated by longer files.

Table 6. Main experimental parameter settings.

Parameter	Original Empirical	Optimized	Final Value
Time window length T	10	10	10
Delay time τ	1	2	2
Embedding dimension m	3	3	3
Power exponent α	0.4	0.4736	0.474
Batch size	64	—	64
Maximum epochs	90	50 *	150
Initial learning rate	0.0005	1.995×10^{-5}	1.995×10^{-5}
Optimizer	Adam	—	Adam
Loss function	Huber loss	—	Huber loss
Early stopping patience	12	8	12
Learning-rate reduction patience	6	6	6
Minimum learning rate	1×10^{-6}	—	1×10^{-6}

* During HPO trials; final training uses 150 epochs with early stopping.

5. Results and Discussion

5.1. Overview of Experimental Analysis

This chapter presents the experimental results and analysis of the proposed MI-PD-TCN model for Valve 1005 opening prediction. The evaluation is conducted from two perspectives. First, comparative experiments are carried out against five baseline models to demonstrate the overall advantage of the proposed method. Second, ablation experiments are performed to quantify the individual contribution of each key component.

All models in the comparative experiments share the same MI-selected feature set to ensure a fair structural comparison. The primary evaluation metrics are MAE, RMSE, and R^2 . MAPE is reported as a supplementary reference only, because near-zero valve opening values cause MAPE to become extremely large and numerically unstable.

5.2. Comparative Experiment Results

To rigorously validate the proposed MI-PD-TCN model and address concerns about statistical significance, we conducted 10 independent runs using different random seeds. All experiments employed identical train-validation-test splits and shared the same test dataset, ensuring fair comparison across all baseline methods. This multi-run protocol provides reliable mean values, standard deviations, and enables statistical hypothesis testing.

Table 7 presents comprehensive performance statistics for all six algorithms. The results demonstrate that the proposed MI-PD-TCN consistently outperforms baseline methods across all three primary metrics. The mean R^2 reaches 0.9552 ± 0.0062 , representing the highest performance among all evaluated methods. Notably, the standard deviation of MI-PD-TCN's R^2 (± 0.0062) is substantially smaller than its competitors, indicating superior stability and robustness to random initialization.

Among the deep learning baselines, LSTM represents the strongest competitor with $R^2 = 0.9342 \pm 0.0073$. Comparison between MI-PD-TCN and LSTM reveals statistically

significant improvements across all metrics. The RMSE reduction from 3.863 to 3.199 represents a 17.2% improvement, with a Cohen's *d* effect size of 2.75 ($p < 0.01$), indicating a large and highly significant effect. The R^2 improvement of 2.25 percentage points (from 0.9342 to 0.9552) is accompanied by an effect size of $d = 2.91$ ($p < 0.01$), further confirming the statistical significance of the proposed method.

Table 7. Comparative prediction results of different models for Valve 1005.

Algorithm	MAE ($\mu \pm \sigma$)	RMSE ($\mu \pm \sigma$)	R^2 ($\mu \pm \sigma$)	CV (R^2) *
XGBoost	4.412 ± 0.068	5.571 ± 0.043	0.8622 ± 0.0021	0.24%
LightGBM	3.738 ± 0.145	5.849 ± 0.067	0.8492 ± 0.0040	0.47%
LSTM	2.231 ± 0.131	3.863 ± 0.241	0.9342 ± 0.0073	0.78%
CNN	2.659 ± 0.668	4.252 ± 0.482	0.9220 ± 0.0193	2.10%
TCN Basic	2.862 ± 0.401	3.840 ± 0.471	0.9357 ± 0.0128	1.37%
MI-PD-TCN	2.080 ± 0.242	3.199 ± 0.233	0.9552 ± 0.0062	0.65%

* CV (Coefficient of Variation) measures relative volatility of R^2 across runs; lower is better.

Beyond accuracy, a critical distinction emerges in model stability. The coefficient of variation for MI-PD-TCN's R^2 is 0.65%, substantially lower than LSTM's 0.78% and competitive with tree-based ensemble methods (XGBoost 0.24%, LightGBM 0.47%). This indicates that MI-PD-TCN maintains consistent performance across different initializations, a desirable property for production deployment. In contrast, CNN exhibits the highest variability (CV = 2.10%) among neural network methods, suggesting that convolutional architectures are more sensitive to random seeds in this application domain.

The stability advantage becomes evident when examining the range of performance across all 10 runs. LSTM's R^2 ranges from 0.9223 to 0.9443 (span of 0.0220), whereas MI-PD-TCN's range is 0.9465 to 0.9639 (span of 0.0174). More importantly, even the worst-case performance of MI-PD-TCN ($R^2 = 0.9465$ at No.2) exceeds the best-case performance of LSTM ($R^2 = 0.9443$ at No.6). This observation strongly suggests that the performance improvement stems from fundamental algorithmic superiority rather than variance or initialization luck, lending credibility to the proposed approach.

The detailed per-seed results in Table 8 further illustrate the consistency of MI-PD-TCN across different random initializations. While LSTM exhibits substantial variation in RMSE (3.548 to 4.191), MI-PD-TCN's RMSE ranges from 2.856 to 3.479, demonstrating the distinct performance characteristics of the proposed architecture when applied to this wind tunnel system. The tightly constrained R^2 range of MI-PD-TCN across all 10 seeds reflects the stabilizing effect of mutual information-guided feature selection and progressive dilation mechanisms. This consistency is particularly valuable in practical applications where model reliability is paramount.

Table 8. Detailed RMSE Performance Across Random Seeds.

Seed No.	LSTM RMSE	MI-PD-TCN RMSE	Improvement
1	4.191	3.316	−20.9%
2	3.739	3.479	−6.9%
3	3.667	3.132	−14.6%
4	3.918	2.948	−24.8%
5	4.148	3.225	−22.3%
6	3.548	2.856	−19.5%

Table 8. *Cont.*

Seed No.	LSTM RMSE	MI-PD-TCN RMSE	Improvement
7	3.727	3.220	−13.6%
8	3.845	3.388	−11.9%
9	3.846	2.877	−25.2%
10	3.842	2.951	−23.2%

5.3. Ablation Experiment Design

To quantify the contribution of each component, ablation experiments were designed as summarized in Table 9. The complete model is denoted as E0. Most experiments remove or replace one specific component while keeping all others consistent with E0.

Table 9. Ablation experiment settings.

ID	Setting	Description
E0	Complete MI-PD-TCN	Full model with all proposed components
E1	Without MI selection	All candidate features used without MI screening
E2	Without power transformation	Target power transformation removed
E3	Without PSR	Phase space reconstruction removed
E4	Standard TCN baseline	Simplified standard TCN; MI selection, power transform, PSR, progressive dilation, sigmoid output, and sample weighting all removed
E5	Without progressive dilation	Progressive dilation disabled; all other components retained
E6	Pre-Norm structure	Post-Norm replaced by Pre-Norm
E7	Without sigmoid output	Sigmoid output constraint removed
E8	Without sample weighting	Equal sample weights for all training files

Experiment E4 is an exception. It is designed as a simplified standard TCN baseline in which multiple enhancement modules are removed simultaneously, including MI feature selection, power transformation, PSR, progressive dilation, sigmoid output constraint, and sample weighting. Therefore, E4 represents a framework-level comparison against a conventional TCN setting, rather than a single-factor ablation. Its performance degradation reflects the combined effect of removing all these components and cannot be attributed to any individual module.

5.4. Ablation Experiment Results

Each ablation setting was run multiple times to assess both accuracy and stability. Table 10 and Figure 6 reports the best run and the worst run for each experiment, selected by lowest and highest RMSE respectively. This unified two-row format provides a consistent basis for stability comparison across all settings. Figures 7 and 8 gives the average R2 comparison and Best-worst RMSE interval plot for stability comparison across E0–E8 respectively.

Table 10. Average performance summary of ablation experiments.

ID	MAE (Avg)	RMSE (Avg)	R ² (Aavg)	Key Observation
E0	2.2681	3.1932	0.9537	Best accuracy and stability
E1	2.7771	4.1440	0.9181	Large best-worst gap; unstable

Table 10. Cont.

ID	MAE (Avg)	RMSE (Avg)	R^2 (Avg)	Key Observation
E2	3.7184	4.8323	0.8940	Largest single-factor accuracy drop
E3	2.6566	3.6851	0.9386	Moderate accuracy decrease
E4	5.5115	8.4855	0.6768	Severe degradation; framework baseline
E5	2.8461	4.0583	0.9265	Accuracy drop without dilation
E6	2.2324	3.4170	0.9473	Minor difference from E0
E7	4.0555	5.2808	0.8754	Output constraint is critical
E8	3.4209	5.1352	0.8603	High instability across runs

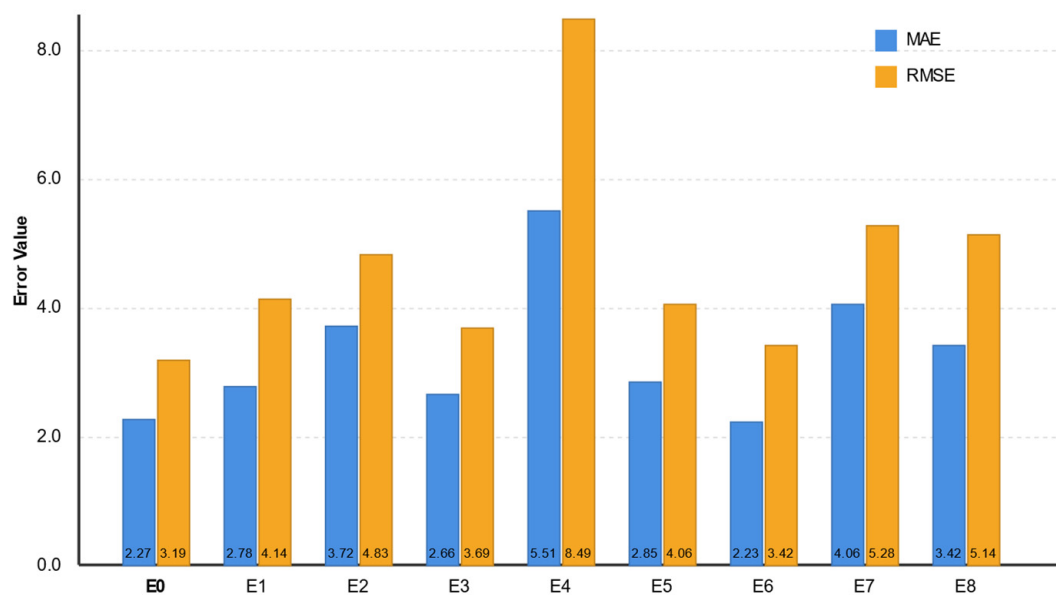
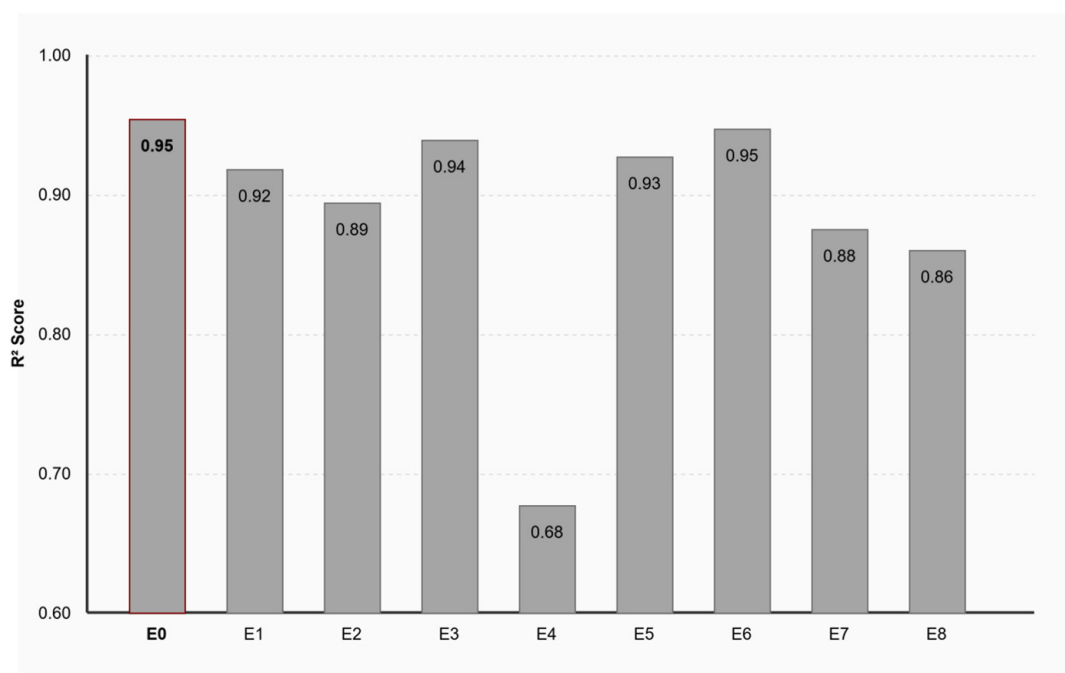


Figure 6. Grouped bar chart comparing average MAE and RMSE across ablation experiments E0–E8.

Figure 7. Bar chart comparing average R^2 across ablation experiments E0–E8.

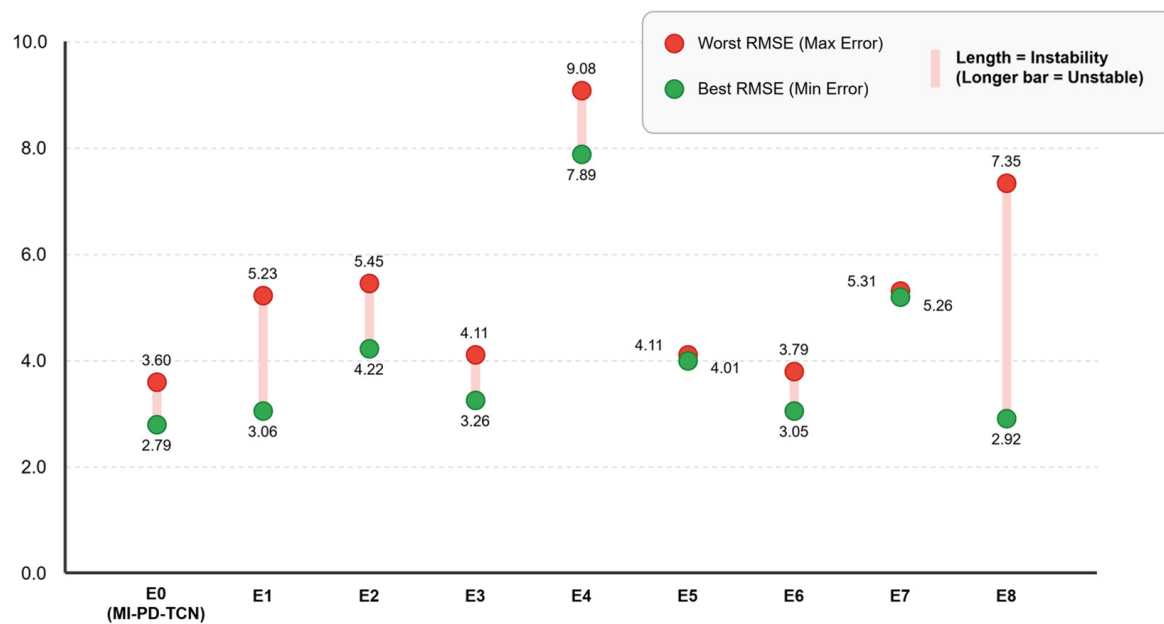


Figure 8. Best-worst RMSE interval plot for stability comparison across E0–E8.

5.5. Analysis of Ablation Results

5.5.1. Effect of Multi-Target Mutual Information Feature Selection

Experiment E1 replaces MI-selected features with all candidate features. The average RMSE increases from 3.1932 to 4.1440 and average R^2 decreases from 0.9537 to 0.9181. More importantly, the best-worst RMSE gap of E1 ($5.2253 - 3.0626 = 2.1627$) is considerably larger than that of E0 ($3.6007 - 2.7857 = 0.8150$), indicating significantly reduced training stability.

Using all candidate variables without MI screening introduces redundant and weakly relevant industrial signals. Although deep learning models possess automatic feature learning capacity, excessive irrelevant inputs increase optimization difficulty and amplify sensitivity to random initialization. The multi-target MI criterion, which evaluates feature relevance jointly against multiple coupled valve targets, effectively reduces input dimensionality while retaining the most informative variables, thereby improving both accuracy and training robustness.

5.5.2. Effect of Power Transformation

Experiment E2 removes the target power transformation. The average RMSE increases from 3.1932 to 4.8323 and average R^2 decreases from 0.9537 to 0.8940, representing the largest accuracy drop among all single-factor ablation experiments.

As described in Section 3.4, the power transformation compresses large target values and reduces the skewness of the valve opening distribution. Without this transformation, the raw target distribution retains a heavy tail and large fluctuations, making the regression objective harder to optimize. The result confirms that target distribution adjustment is a critical preprocessing step for valve opening prediction.

5.5.3. Effect of Phase Space Reconstruction

Experiment E3 removes phase space reconstruction. The average RMSE increases from 3.1932 to 3.6851 and average R^2 decreases from 0.9537 to 0.9386.

As described in Section 3.5, PSR constructs an augmented input vector by embedding delayed historical observations alongside the current feature vector. For wind tunnel valve control, the current opening state is influenced not only by instantaneous sensor readings

but also by the recent dynamic evolution of the system. Removing PSR deprives the model of this historical context, weakening its temporal representation capacity and leading to a moderate performance decrease.

5.5.4. Comparison with the Simplified Standard TCN Baseline

Experiment E4 provides a framework-level comparison by replacing the complete MI-PD-TCN pipeline with a simplified standard TCN in which multiple enhancement modules are removed simultaneously. The average RMSE increases sharply from 3.1932 to 8.4855 and average R^2 decreases from 0.9537 to 0.6768. Even the best run of E4 (RMSE = 7.8876, $R^2 = 0.7221$) is far inferior to the worst run of E0 (RMSE = 3.6007, $R^2 = 0.9421$).

Since MI selection, power transformation, PSR, progressive dilation, sigmoid output, and sample weighting are all removed in E4, the observed degradation reflects the combined absence of all these components. The result should not be interpreted as the isolated effect of the progressive dilation structure. The individual contribution of progressive dilation is analyzed separately in Section 5.5.5 using the single-factor ablation E5. The comparison between E4 and E0 collectively demonstrates the strong overall advantage of the complete MI-PD-TCN framework over a conventional standard TCN setting.

5.5.5. Effect of Progressive Dilation

Experiment E5 disables only the progressive dilation mechanism while retaining all other components of the complete model. The average RMSE increases from 3.1932 to 4.0583 and average R^2 decreases from 0.9537 to 0.9265.

This single-factor comparison confirms that the progressive dilation structure itself provides a meaningful accuracy improvement. As described in Section 3.6, the progressive dilation strategy expands the temporal receptive field in a structured manner, enabling the model to simultaneously capture short-term fluctuations and longer-range dependencies in the valve opening sequence. Disabling this mechanism reduces temporal modeling capacity and leads to a measurable accuracy drop, confirming the value of the progressive dilation design.

5.5.6. Effect of Normalization Position

Experiment E6 replaces Post-Norm with Pre-Norm in the residual blocks. The average RMSE increases slightly from 3.1932 to 3.4170 and average R^2 decreases from 0.9537 to 0.9473. The best-worst RMSE gap of E6 ($3.7865 - 3.0475 = 0.7390$) is close to that of E0 (0.8150), indicating comparable stability.

The result shows that the model is not highly sensitive to normalization position. However, since Post-Norm consistently yields slightly better average performance, it is retained in the final MI-PD-TCN structure. E6 is regarded as a minor structural variant rather than a critical ablation component.

5.5.7. Effect of Sigmoid Output Constraint

Experiment E7 removes the sigmoid activation function from the output layer. The average RMSE increases from 3.1932 to 5.2808 and average R^2 decreases from 0.9537 to 0.8754. Furthermore, the best-worst gap of E7 is small ($5.3063 - 5.2553 = 0.0510$), but consistently poor, indicating that removing the sigmoid output causes systematic rather than random degradation.

The sigmoid constraint confines predictions to the range $[0, 1]$ to match the normalized target domain. Without this constraint, the model may generate out-of-range predictions, particularly during rapid valve transitions. These violations are amplified upon inverse transformation back to the original scale, producing systematically larger errors. The result

confirms that the sigmoid output constraint is a critical component for numerically stable valve opening prediction.

5.5.8. Effect of Sample Weighting

Experiment E8 removes file-level sample weighting and assigns equal weights to all training samples. The average RMSE increases from 3.1932 to 5.1352, average R^2 decreases from 0.9537 to 0.8603, and the best-worst RMSE gap is $7.3505 - 2.9199 = 4.4306$, which is the largest instability observed among all ablation experiments.

The training dataset consists of multiple experimental files with different sequence lengths. Without weighting, longer files contribute proportionally more samples and may dominate the loss function, causing the model to overfit to certain operating conditions while under performing on shorter sequences. The file-level weighting strategy, as described in Section 3.7 normalizes the effective contribution of each file regardless of its length. The ablation result confirms that this strategy is essential not only for accuracy but also for cross-condition robustness.

5.6. Stability Analysis

The best-worst RMSE gaps reported in Table 9 provide a direct measure of training stability across runs. The key observations are as follows.

E0 achieves both the best average performance and one of the smallest best-worst gaps (0.8150), confirming that the complete framework is both accurate and stable. E1 and E8 exhibit the largest fluctuations, with best-worst RMSE gaps of 2.1627 and 4.4306 respectively, confirming that MI feature selection and sample weighting are the two components most critical for training stability. E2 and E4 both show persistently high RMSE in both best and worst runs, indicating systematic rather than random degradation caused by removing power transformation or replacing the full framework.

These stability results supplement the average accuracy comparison and collectively support the design of the complete MI-PD-TCN model.

5.7. Discussion

The comparative and ablation experiments lead to the following key conclusions.

Multi-target MI feature selection improves both accuracy and stability. It reduces average RMSE by 22.9% compared with E1 and significantly narrows the best-worst performance gap.

Power transformation is the most impactful single preprocessing step. Removing it increases average RMSE by 51.4% compared with E0, confirming that target distribution adjustment is essential for the uneven valve opening data.

Phase space reconstruction provides moderate but consistent improvement. It enhances temporal context without introducing structural complexity, and its removal leads to a measurable decrease in R^2 .

Progressive dilation (E5 vs. E0) provides a clear single-factor accuracy gain. The framework-level comparison with E4 further demonstrates that the complete pipeline significantly outperforms a simplified standard TCN, though the E4 result itself reflects the combined absence of multiple components.

Sigmoid output constraint produces systematic improvement by enforcing physically meaningful prediction bounds. Its removal leads to consistently poor performance across all runs of E7.

Sample weighting is the most critical factor for training stability. Its removal in E8 causes the largest best-worst RMSE gap among all ablation experiments, confirming its importance for robust cross-condition training.

Overall, the proposed MI-PD-TCN model achieves the best prediction accuracy and training stability through the coordinated design of feature selection, target preprocessing, temporal reconstruction, progressive dilation, output constraint, and sample weighting. Each component contributes measurably to the final performance, and no single module alone accounts for the full advantage of the proposed framework.

6. Conclusions

This study proposed a mutual information-based progressive dilated temporal convolutional network, named MI-PD-TCN, for valve opening prediction in a wind tunnel control system. The proposed method integrates data-driven feature selection, nonlinear temporal state reconstruction, and deep temporal feature extraction into a unified predictive framework. First, multi-source physical sensing signals and valve-related operational variables were preprocessed to construct candidate features, including historical lag features and differential features. Then, a multi-target mutual information selection strategy was introduced to identify informative variables associated with different valve opening targets, thereby reducing redundant inputs and improving feature relevance. After feature normalization and phase space reconstruction, sliding-window temporal tensors were generated as the model input. Finally, a progressive dilated TCN backbone was designed to extract multi-scale temporal dependencies through gradually enlarged receptive fields.

The experimental results demonstrate that the proposed MI-PD-TCN model can effectively capture the nonlinear dynamic relationship between multi-source sensing variables and target valve opening sequences. Compared with conventional baseline models, the proposed model achieves higher prediction accuracy and more stable generalization performance. The mutual information-based feature selection module contributes to a more compact and discriminative input representation, while the phase space reconstruction and sliding-window strategy enhance the temporal structure of the input data. In addition, the progressive dilated convolution design enables the model to learn both short-term fluctuations and long-range dependencies without excessively increasing model complexity.

Overall, the proposed framework provides an effective solution for data-driven valve opening prediction under complex wind tunnel operating conditions. It improves the interpretability of feature selection and strengthens the temporal modeling capability of the prediction network. The method can also be extended to other industrial control scenarios involving multi-source sensor data and nonlinear actuator behavior.

However, this study still has some limitations. The current model mainly relies on offline historical data and has not yet been fully validated under real-time closed-loop control conditions. In addition, the selection of hyper parameters, such as the sliding-window length, embedding dimension, and dilation configuration, may influence prediction performance under different operating conditions. Future work will focus on online deployment, adaptive hyper parameter optimization, uncertainty estimation, and the incorporation of physical constraints to further improve the robustness and reliability of valve opening prediction in practical wind tunnel control systems.

Author Contributions: Conceptualization, Q.L. and Y.X.; methodology, M.C.; software, L.C. and M.C.; formal analysis, D.L.; investigation, Y.X.; resources, J.G.; writing—original draft preparation, Q.D.; visualization, Y.X.; supervision, L.C.; project administration, J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data generated in this study cannot be publicly released due to privacy and ethical restrictions. Reasonable data access requests can be directed to the corresponding author upon valid application.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yi, F.; Zhong, R.X.; Zhu, W.J.; Zhou, R.; Wang, Y.; Guo, L. Leakage fault diagnosis of wind tunnel valves using wavelet packet analysis and vision transformer-based deep learning. *Mathematics* **2025**, *13*, 3195. [\[CrossRef\]](#)
2. Zheng, R.; Lei, L. Multiscale physics-informed modeling and hybrid intelligent control of urban district heating networks. *Energy* **2026**, *344*, 140117. [\[CrossRef\]](#)
3. Zhao, Z.; Yi, X.; Xiong, J.J.; Ran, L.; Zhao, J.Y.; Zhu, Y.L. AAF-Bi-LSTM-NARX: A bidirectional network with adaptive attention fusion for sensor data imputation. *Expert Syst. Appl.* **2026**, *310*, 131266. [\[CrossRef\]](#)
4. Han, Z.Y.; Zhao, J.; Leung, H.; Ma, A.; Wang, W. A review of deep learning models for time series prediction. *IEEE Sens. J.* **2021**, *21*, 7833–7848. [\[CrossRef\]](#)
5. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, arXiv:1803.01271. [\[CrossRef\]](#)
6. Zhang, L.; Ren, G.F.; Li, S.L.; Du, J.S.; Xu, D.Y.; Li, Y.H. A novel soft sensor approach for industrial quality prediction based TCN with spatial and temporal attention. *Chemom. Intell. Lab. Syst.* **2025**, *257*, 105272. [\[CrossRef\]](#)
7. Yuan, C.Q.; Xie, Y.F.; Xie, S.W.; Wang, J. Pruned tree-structured temporal convolutional networks for quality variable prediction of industrial process. *J. Process Control* **2024**, *143*, 103312. [\[CrossRef\]](#)
8. Yao, Q.L.; Zhao, X.Q.; Huo, H.F.; Hui, Y.Y. Quality prediction of industrial process based on Kolmogorov-Arnold graph convolution aggregation temporal convolution network. *Meas. Sci. Technol.* **2025**, *36*, 066207. [\[CrossRef\]](#)
9. Cortés-Ibáñez, J.A.; González, S.; Valle-Alonso, J.J.; Luengo, J.; García, S.; Herrera, F. Preprocessing methodology for time series: An industrial world application case study. *Inf. Sci.* **2020**, *514*, 385–401. [\[CrossRef\]](#)
10. Pan, H.; Zhang, P.J.; Xue, C.; Liu, D. Interpretable data-driven crystal diameter prediction in CZ silicon single-crystal growth via MIC-guided and GWO-optimized TCN-LSTM. *Processes* **2026**, *14*, 1153. [\[CrossRef\]](#)
11. Lima, R.D.; Scardua, L.A.; de Almeida, G.M. Predicting temperatures inside a steel slab reheating furnace using neural networks. *IEEE Trans. Ind. Appl.* **2025**, *61*, 5273–5282. [\[CrossRef\]](#)
12. Zhou, K.; Wang, W.Y.; Hu, T.; Deng, K. Time series forecasting and classification models based on recurrent with attention mechanism and generative adversarial networks. *Sensors* **2020**, *20*, 7211. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Li, G.N.; Chen, J.J.; Tao, Y.T.; Li, A.X.; Zhang, X.W. Residual life prediction of pneumatic control valves based on Trans-TCN-GRU modeling. *IEEE Access* **2024**, *12*, 191670–191681. [\[CrossRef\]](#)
14. Wang, H.; Peng, M.J.; Xu, R.Y.; Ayodeji, A.; Xia, H. Remaining useful life prediction based on improved temporal convolutional network for nuclear power plant valves. *Front. Energy Res.* **2020**, *8*, 584463. [\[CrossRef\]](#)
15. Gao, S.W.; Li, X.; Yang, W.B.; Xie, J.J.; Yun, P.X. Industrial prediction method based on graph sampling and aggregation of temporal features. *Can. J. Chem. Eng.* **2026**, *104*, 2083–2103. [\[CrossRef\]](#)
16. Tuo, B.B.; Zhao, X.Q.; Sun, K.W.; Liu, K.; Hui, Y.Y. Soft sensor model for nonlinear dynamic industrial process based on GraphSAGE-IMATCN. *Process Saf. Environ. Prot.* **2024**, *191*, 1131–1147. [\[CrossRef\]](#)
17. Han, M.; Zhong, K.; Qiu, T.; Han, B. Interval type-2 fuzzy neural networks for chaotic time series prediction: A concise overview. *IEEE Trans. Cybern.* **2019**, *49*, 2720–2731. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Pan, C.Q.; Su, L.Y.; Li, Q.S.; Chen, X.J.; Li, F.L. MSK-TCN: Multi-scale stacked kernel temporal convolutional network for multivariate chaotic time series prediction. *Phys. Scr.* **2025**, *100*, 065211. [\[CrossRef\]](#)
19. Su, Y.; Yang, C.L.; Qiao, J.F. Effluent ammonia nitrogen prediction using a phase space reconstruction method combining pipelined recurrent wavelet neural network. *Appl. Soft Comput.* **2022**, *120*, 108602. [\[CrossRef\]](#)
20. Huang, W.J.; Li, Y.T.; Huang, Y. Deep hybrid neural network and improved differential neuroevolution for chaotic time series prediction. *IEEE Access* **2020**, *8*, 159552–159565. [\[CrossRef\]](#)
21. Cheng, J.Z.; Huang, K.Z.; Zheng, Z.B. Fitting imbalanced uncertainties in multi-output time series forecasting. *ACM Trans. Knowl. Discov. From Data* **2023**, *17*, 102. [\[CrossRef\]](#)
22. Hssayeni, M.D.; Ghoraani, B. Deep regression modeling for imbalanced and incomplete time-series data. *IEEE Trans. Emerg. Top. Comput. Intell.* **2024**, *8*, 3767–3778. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.