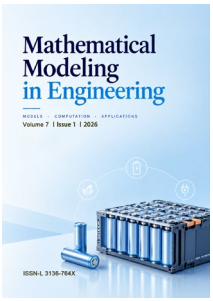
UNIVERSITY OF WEST ATTICA

eISSN: 3136-764X

Mathematical Modeling in Engineering

Vol. 7, No. 2, April, 2026



STID-Mixer Model for Polymer Melt Index Prediction from Extrusion Process Sensor Streams

Hrvoja Rajšić[a,*], Kraljevka Novčić[a], Ignjata Milutinović[a]

[a] *Faculty of Computational Chemical Engineering, Union University, Belgrade, 11000, Serbia*

Received 07 January 2026; Revised 09 April 2026; Accepted 13 April 2026; Published 17 April 2026

Abstract

Online estimation of polymer melt index is difficult because laboratory assays become available only after the corresponding material has progressed beyond the critical extrusion stages, whereas plant historians continuously record thermal, pressure, torque, speed, and feeding behavior. This study developed STID-Mixer, a compact soft sensor that converts multirate extrusion streams into a residence-aware representation for continuous melt-index prediction. Historian signals were synchronized with laboratory samples, screened by production state, normalized within product grade, and transformed into sensor-time tokens containing dynamic descriptors and material-residence weights. The network then separated cross-sensor interaction learning from temporal-delay mixing and used a robust regression objective. Evaluation used 18,420 one-minute windows derived from 17 raw sensor variables, 11 engineered descriptors, and six polymer grades under a chronological campaign-based train-validation-test split. STID-Mixer achieved a test RMSE of 0.118 g/10 min, an MAE of 0.083 g/10 min, an R^2 of 0.936, and 91.7% coverage within ±0.20 g/10 min. Its RMSE was 21.9% lower than that of CatBoost, while single-window inference required 3.4 ms. Ablation, residual, drift, and replay analyses indicated that grade normalization, residence weighting, and separated sensor-time mixing contributed complementary benefits. Errors increased during early grade transitions, upper-range operation, and progressive die-pressure drift. The results support STID-Mixer as a computationally light advisory estimator for extrusion-quality surveillance, while prospective cross-line validation and uncertainty-aware alarms remain necessary before closed-loop use.

Keywords: Polymer Melt Index, Extrusion Sensor Streams, STID-Mixer, Soft Sensor, Process Quality Prediction

1. Introduction

A relatively straightforward index that combines the effects of molecular structure, thermal history and processability for quick reference by production staff is used to quantify the flow properties of thermoplastics [1]. Property measurement of extrusion plants is still generally carried out off-line or at-line; thus, the reported values may only be obtained after the relevant material has gone through the die, cooling, palletization and packing stages [2]. At the same time, the modern lines are also collecting data on barrel zone, adapter and die pressure, screw torque, screw speed, melt temperature, vacuum, feed rate and motor load [3]. Soft sensing can connect the easy-to-obtain measurements with the delayed quality tests, but a good estimator needs to address industrial noise, multicollinearity, transport delay and grade-dependent operating baselines [4]. The need is most prominent during feed variation, thermal overshoot, die fouling and grade transition; at this time, the laboratory result has arrived too late to explain the process abnormality that occurred. The history of the sensor offers an indirect record of alterations in the polymer; it is not a melt index, and torque cannot distinguish between various types of degradation or determine residence time. Thus, the prediction problem needs to build a dynamic link between the latest sensor data and the next laboratory number. Organize the information to show

* Corresponding author: hrvoja.ra@union.edu.rs

https://doi.org/10.68406/mme.2026.vol7iss2nm2:13-26

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

which variables were changed at the same time, when these modifications occurred, and whether the resulting condition was in a stable production or a temporary state of material flow.

Traditionally, melt-index soft sensors employed linear regression, partial least squares or feature-engineered machine-learning models due to their simple calibration and ease of acceptance by plant engineers [5]. The problem is that the extrusion data are not in the form of static tabular samples. Only a large increase in pressure indicates an increase in torque; a change in barrel temperature may be harmless at one screw speed but will affect the quality of the product at another, and feeder instability can lead to quality problems after a certain residence time. Tree ensembles and gradient boosting models can capture non-linear relationships; however, they are also very sensitive to the window statistics selected for training [6]. Although recurrent neural networks and convolutional neural networks reduce the amount of manual feature extraction, they may introduce an order assumption of the sensor variables or be too complex for deployment in a programmable industrial environment [7]. Recently, mixer-style architectures have proposed another way: instead of deep recurrence or attention for each sample, a model can alternate mixing in the feature and time dimensions to learn cross-sensor and delayed-process interactions more economically [8].

This paper addresses the engineering problem of representing coupled extrusion dynamics without requiring a large recurrent or attention-based model. General time-series models can be applied to historian data, but they do not natively encode grade-dependent baselines, material transport delay, or the different roles of rapid pressure disturbances and slow thermal memory. STID-Mixer therefore contributes three linked elements: grade-conditioned robust normalization, a residence-informed temporal prior, and explicit decomposition of sensor interaction mixing from temporal-delay mixing. Each observation window contains aligned one-minute sensor summaries and short-horizon dynamic descriptors, while the residence prior emphasizes positions that plausibly contributed to the sampled material. The contribution is thus an extrusion-specific integration of process representation and lightweight mixer learning, rather than a claim that mixer networks themselves are new. Laboratory assays remain necessary for supervised training and periodic verification; after training, the model produces online melt-index estimates from synchronized historian streams.

The remainder of the paper is organized as follows. Section 2 reviews melt-index soft sensing and temporal learning methods and identifies the unresolved need for an extrusion-specific, interpretable, and deployable representation. Section 3 describes historian alignment, grade normalization, residence-aware tokenization, and deterministic reconstruction of model inputs. Section 4 specifies the STID-Mixer architecture, robust objective, model-selection procedure, and online inference path. Section 5 reports the dataset protocol, baseline comparison, residual analysis, ablation study, sensor-drift stress test, robustness evaluation, and replay experiment. Section 6 states the principal findings, limitations, and requirements for broader industrial validation.

2. Related Work

2.1. Melt Index Soft Sensing

Melt-index prediction is a type of general process soft sensing, and easily obtained parameters are used to infer a relatively slow-changing or difficult-to-measure quality characteristic. In polymer extrusion, due to the nature of sensor variables, they are not independent descriptors but rather combined responses to heating, pressurization and mechanical shearing of the flow. The first generation of industrial soft sensors generally treated the process as a stable regression problem and extracted the mean values of the temperature, pressure, speed and load channels in a certain window [9]. Partial Least Squares is still employed because it can address the issue of multicollinearity and produces coefficients that are more intuitive for process engineers. It is not linear, and at different screw speeds and melt temperatures, the same pressure fluctuation will have different effects.

Nonlinear Learning Methods have improved the situation by representing the interaction of variables without an explicit rheological equation. Support Vector Regression, Random Forest, and Gradient Boosting methods have been applied to predict polymer quality in cases with few laboratory labels and irregular sensor channels [10]. The above models are suitable for medium-sized datasets, but they usually require a hand-crafted feature table. A fixed feature table cannot determine whether the fluctuation of the sensor ten minutes ago is more relevant than the current one for the model. Some scholars have added moving-window statistics, lagged variables and transition flags to the process memory [11]. Although the above descriptors are convenient, they increase the number of features and need to be plant-specifically tuned every time there are changes in equipment geometry, residence time or grade family.

2.2. Time-Series Learning for Extrusion Streams

Time-series models can be employed to directly use the order and persistence of extrusion signals. Recurrent Neural Networks can learn delayed responses via hidden states, and Temporal Convolutional Networks expand the receptive field of past observations using dilated filters [12]. These models have been used to monitor changes in chemical and polymer production by learning from historical data. At present, the applications of these in melt-index prediction are limited by label frequency, grade imbalance and the need for stable deployment. At different times, various materials

have exhibited different laboratory melt index values; therefore, a basic model needs to be established for the placement of sensors and labels to ensure uniformity.

Although attention-based time-series models can better learn long-term dependencies, they are not suitable for small industrial datasets and may hide the relationship between a prediction and known process variables [13]. A single mixer structure can divide the directions of information collection to simplify things. Feature mixing can model cross-sensor coupling, and temporal mixing can model delayed process influence without recurrent state propagation [14]. Separation is convenient for extrusion applications; therefore, engineers can determine the quality based on data from both sets of sensors and identify when these changes will be reflected in the die. Most of the existing mixer models are general-purpose forecasting or classification models and are not suitable for melt-index soft sensing [15]. The Gap model builds a process-aware window, performs interpretable sensor-time mixing, and uses a regression objective suitable for sparse laboratory quality labels.

3. Process Stream Representation

3.1. Historian Alignment and Grade Normalization

Plant historian records were aligned with laboratory melt-index labels before model construction. The retained channels included barrel-zone temperatures, die and adapter pressures, screw torque, screw speed, motor current, melt temperature, feeder rate, vacuum, and ambient correction. Production-state gates based on documented ranges of screw speed, feeder rate, and die-pressure variability excluded start-up, shutdown, purging, and short recipe interruptions. The remaining signals were resampled at one-minute intervals to preserve operational dynamics while attenuating high-frequency variation that could not be resolved by the laboratory assay. All state gates were applied before partitioning so that the same deterministic screening rules were used for training, validation, testing, and online replay.

Let $s_{t,k}$ denote the raw value of sensor k at minute t. For each product grade, a robust baseline and scale are estimated from steady windows rather than from the full historian. This prevents grade transitions and abnormal campaigns from dominating the normalization. The equation below defines the standardized process variable used by the subsequent window construction.

$$x_{t,k} = \frac{s_{t,k} - \mu_{g,k}}{\sigma_{g,k} + \varepsilon} \tag{1}$$

In this equation, $\mu_{g,k}$ and $\sigma_{g,k}$ are the median and robust scale of sensor k for grade g, and ε is a small numerical constant. This normalization leaves the model with deviations from the grade recipe instead of absolute recipe values. It is especially important for barrel-zone temperatures, where a difference of two degrees can be meaningful inside a narrow recipe but irrelevant across unrelated grades.

3.2. Residence-Aware Window Tokenization

Each training sample is a window of L minutes preceding a laboratory label. The window contains standardized sensor values and dynamic descriptors such as local slope, short-term variability, and pressure-to-throughput ratio. The construction is summarized in Figure 1. In this workflow, historian streams enter from the left, steady-state filtering removes unstable intervals, grade normalization centers the retained channels, residence-aware tokenization forms sensor-time patches, and the resulting tokens are sent to the STID-Mixer training and online inference branches.

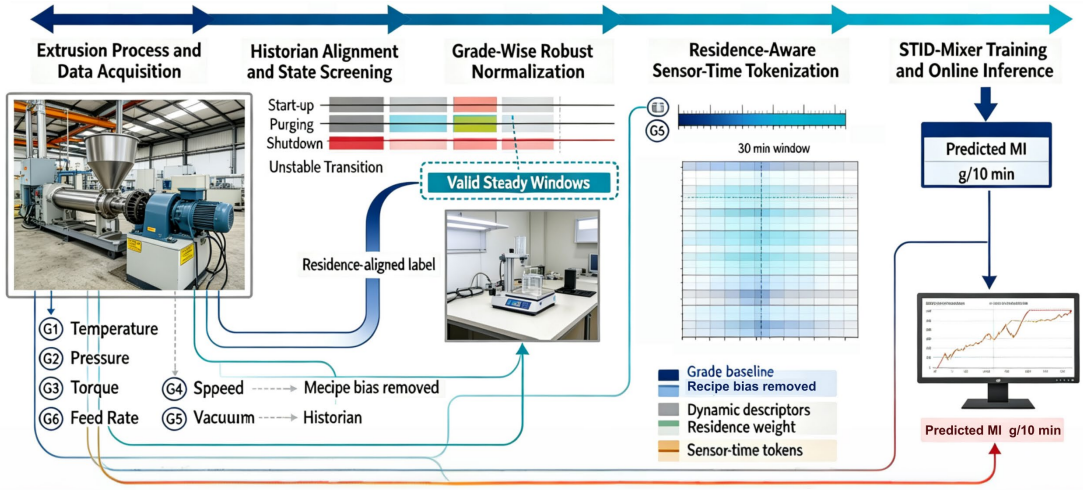


Figure 1. Workflow of extrusion historian alignment, steady-state filtering, grade normalization, residence-aware tokenization, and STID-Mixer inference

A residence-aware temporal weight is introduced because material sampled at the die is influenced by a range of upstream times rather than by the most recent minute alone. The center and width of the residence mask are estimated from screw speed, feed rate, and nominal line geometry. Rather than imposing a rigid transport law, the following equation supplies a smooth prior that emphasizes plausible process-memory positions within the window.

$$a_t = exp\left[-\frac{\left(t-\tau_g\right)^2}{2\rho_g^2}\right], X_t = a_t[x_t, \Delta x_t, v_t] \tag{2}$$

The weighting equation produces a model input tensor that contains sensor channels, engineered dynamic descriptors, and residence weights. For a window length L and C process channels after descriptor expansion, the token matrix X belongs to $\mathbb{R}^{L\times C}$. The label y is the laboratory melt index aligned to the material interval represented by the window. This alignment strategy keeps the method compatible with ordinary plant data, where the laboratory value is delayed but the historian is continuous.

The representation also includes safeguards against two common sources of industrial bias. The first is tag redundancy. Several temperature zones may move together during routine heating control, and several pressure tags may react to the same die restriction. Instead of deleting these channels before modeling, the workflow retains them after robust scaling so that the mixer can learn whether redundancy is harmless or informative under a specific grade. The second source is operator intervention. Manual changes in screw speed or feeder setting are not treated as noise; they are retained when the line has returned to a stable interval, because the model must learn quality behavior under permitted operating adjustments rather than under an idealized recipe.

The windowing procedure was chosen to avoid a frequent mismatch between laboratory sampling and historian timestamps. A laboratory melt-index result normally corresponds to material collected over a short physical interval, yet the historian reports a continuous stream of process states. The sample is therefore assigned to the sensor window whose residence-weighted center best matches the estimated material path. This rule is simpler than a full mechanistic transport model, but it prevents the label from being paired only with the timestamp at which the laboratory result was entered. In practical deployment, the same rule can be implemented with historian timestamps, sample IDs, and a grade-dependent residence-time table maintained by process engineers.

Another practical design choice is to keep the preprocessing deterministic. Every value used by the model can be regenerated from the historian, the grade code, and the laboratory timestamp. This is important for engineering review because a melt-index estimate may later be compared with a retained sample, an operator log, or a customer quality record. The workflow therefore avoids random data augmentation and avoids manual removal of isolated points unless the removal is triggered by a documented rule. Deterministic preprocessing also makes model retraining auditable: if new laboratory labels are added, the same windows can be rebuilt and compared with the previous training set without ambiguity.

4. STID-Mixer Architecture and Training Objective

4.1. Sensor and Temporal Interaction Mixing

STID-Mixer separates cross-sensor interaction learning from temporal-delay aggregation. The 28-channel token matrix is projected to a 96-dimensional latent state and processed by two pre-normalized residual mixer blocks. Within each block, the sensor mixer acts across channel descriptors at every temporal position, whereas the temporal mixer acts across the 30-minute window and receives the residence prior. A gated fusion layer combines the two paths before a scalar regression head produces the melt-index estimate. No spatial ordering is imposed on temperature, pressure, torque, speed, or feeding channels. This compact block order preserves direct residual paths and avoids the sequential computation of a recurrent stack while retaining an explicit location for process-memory information.

The first block projects the token matrix into a latent process state. Sensor interaction decomposition is performed by applying a channel-mixing operator to each temporal position. This operation lets the model learn relations such as pressure-speed coupling, torque-temperature compensation, and feeder-load imbalance. The corresponding sensor-mixing equation is given below.

$$H_s = \phi(XW_s + B_s) + XW_r \tag{3}$$

Within the sensor-mixing equation, W_s maps sensor descriptors into a hidden interaction space, ϕ is a GELU nonlinearity, and B_s is a bias term. A residual connection carries the original standardized variables into the next block, which limits the risk of losing simple linear effects. Temporal-delay mixing is then applied along the window dimension. Unlike recurrence, this step aggregates all positions in the residence window through a learned mixing matrix while retaining the residence mask as an additive prior. The temporal update is captured by the next equation.

$$H_t = \phi(W_t^{\top} H_s + M_r) + H_s \tag{4}$$

The temporal equation combines the learned mixing matrix W_t, the sensor-mixed state H_s, and the residence prior M_r while retaining a residual path. The architecture is shown in Figure 2. Sensor-time tokens first pass through a sensor mixer, then a temporal-delay mixer, then a gated fusion layer, and finally a regression head. The online branch uses the same preprocessing path as the training branch, so deployment does not require a separate feature generator beyond the historian alignment and grade-normalization modules.

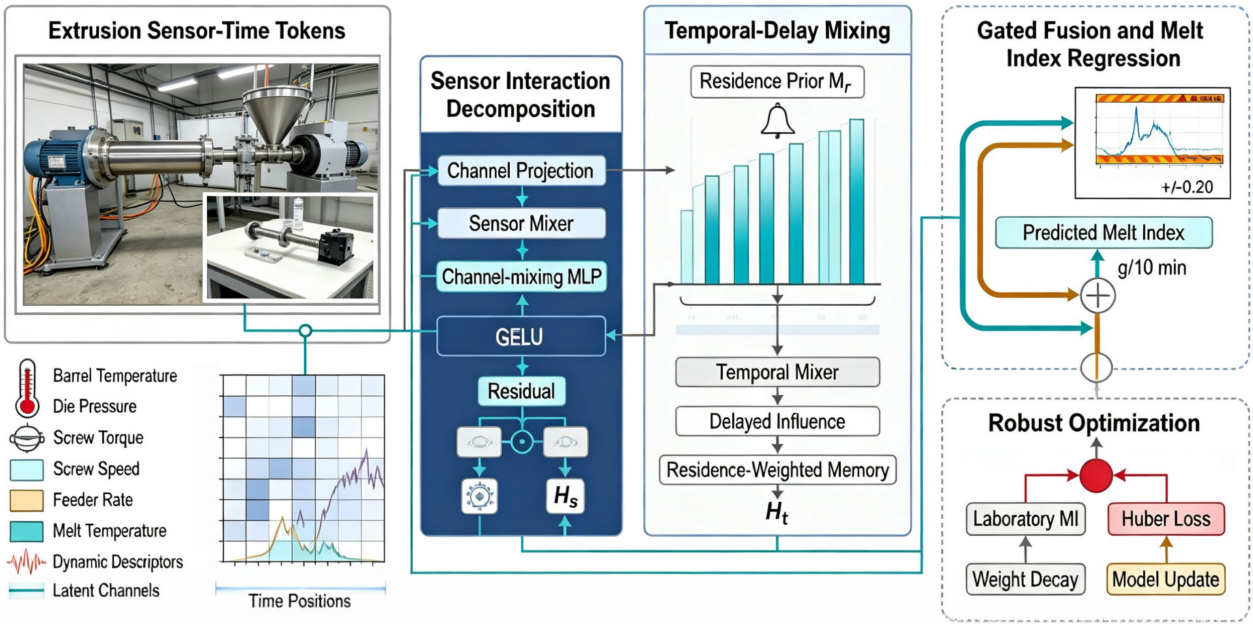


Figure 2. STID-Mixer architecture with sensor interaction decomposition, temporal-delay mixing, gated fusion, and robust melt-index regression

4.2. Robust Optimization and Online Inference

The prediction head maps the fused hidden state to a scalar melt-index estimate. Because laboratory labels can contain sampling noise and because grade transitions may introduce uncertain alignment, training uses a Huber objective rather than pure squared error. The complete loss equation also includes a small weight-decay term to keep the mixer matrices stable.

$$L = \frac{1}{N}\sum_{i} Huber_{\delta}(y_i - \hat{y}_i) + \lambda \|\Theta\|_2^2 \tag{5}$$

This optimization equation balances resistance to isolated label deviations with regularization of the trainable parameter set Θ. The Huber threshold controls the transition between quadratic and linear penalty behavior, whereas λ determines the strength of weight decay. Both quantities are chosen on the validation period rather than adjusted from test residuals.

Model selection used a grade-stratified validation period that followed the training campaigns chronologically. Candidate window length, hidden dimension, dropout rate, Huber threshold, and block depth were selected by validation RMSE together with inspection of the central residual interval; the held-out test campaigns were evaluated only after the configuration was fixed. The selected model used a 30-minute window, a 96-dimensional hidden state, two mixer blocks, and dropout of 0.10. Early stopping rejected configurations whose small RMSE improvement was accompanied by a wider residual tail. These rules document the selection logic available from the experiment record; a complete optimizer and random-seed log should be archived in future releases to support exact computational replication.

The sensor mixer and temporal mixer are deliberately kept shallow. In preliminary trials, deeper stacks reduced training loss but gave less stable validation residuals during grade transitions. The selected two-block design provides one block for primary pressure, temperature, torque, and feed interactions and a second block for delayed corrections. Layer normalization is applied before each mixing operation so that channels with different engineering units do not dominate the latent state after projection. Dropout is applied only inside the mixer blocks, not to the final regression head, because the head must preserve small melt-index differences that are meaningful for quality control.

For online inference, the model receives the newest stable process window and returns a scalar estimate. The preprocessing path caches grade baselines and updates dynamic descriptors as new historian samples arrive. If a required

tag is missing for a short interval, the last valid standardized value is carried forward only within a predefined hold time; longer gaps mark the window invalid. This conservative rule avoids producing a confident melt-index estimate from incomplete process evidence. The resulting pipeline is therefore a soft sensor rather than an autonomous controller: it supplies a continuously updated quality estimate that can be reviewed alongside laboratory assays, recipe limits, and operator knowledge.

The training procedure uses mini-batches stratified by grade so that common grades do not dominate every optimization step. Within each grade, windows remain chronologically ordered before batching, which prevents leakage from future standardization statistics. Validation is performed on later campaigns rather than on a random subset of the training period. The early-stopping criterion monitors both validation RMSE and the width of the central residual interval. A candidate model is rejected if it achieves a slightly lower RMSE by increasing the residual tail, because large tail errors are more damaging for quality monitoring than a small change in average error.

5. Experimental Data and Model Analysis

5.1. Experimental Protocol and Benchmark Accuracy

A continuous extrusion historian was aligned with laboratory melt-index tests to construct the experimental dataset. Each quality result was linked to the upstream process interval that produced the sampled material rather than to the time at which the laboratory result was entered [16]. After start-up, shutdown, purge, and invalid-tag periods were removed, 18,420 labelled one-minute windows remained across six product grades. Each window contained 17 raw sensors and 11 derived descriptors, giving 28 channels before latent projection. Complete production campaigns were assigned chronologically to training, validation, and testing in proportions of 65%, 15%, and 20%, respectively; windows belonging to the same campaign were not distributed across partitions. This campaign-level rule reduced leakage from highly correlated neighboring windows [17]. The test partition contained 3,684 windows, with melt index from 0.72 to 4.86 g/10 min, die pressure from 8.4 to 16.2 MPa, and screw speed from 76 to 124 r/min. Grade-specific counts ranged from 2,610 to 3,420 windows in the complete dataset. Missing-value screening removed 1.8% of candidate windows because of feeder or vacuum-tag interruptions, and no target value was imputed. Grade-wise normalization statistics were fitted on the training period and frozen for validation and testing. Transition windows were retained because continuous quality monitoring is most valuable under changing conditions; consequently, the reported test set was not restricted to an artificially simple steady-state subset.

The benchmark included partial least squares, support vector regression, random forest, CatBoost, LSTM, and a temporal convolutional network, representing linear latent-variable, kernel, tree-ensemble, recurrent, and convolutional approaches to process soft sensing [18]. Every method used the same chronological partitions and the same 30-minute source interval. Neural models received ordered windows, whereas non-neural models received means, standard deviations, slopes, extrema, and residence-weighted summaries computed only from that interval. Hyperparameters were selected on the validation campaigns using fixed candidate grids. RMSE, MAE, R^2, and the fraction of estimates within ±0.20 g/10 min were reported because complementary error and tolerance measures are needed for operational interpretation [19]. Latency was measured after a warm-up pass with batch size one on the same evaluation workstation. The table reports point estimates from the fixed held-out split rather than confidence intervals from repeated independent campaigns; accordingly, comparisons are stated as observed performance differences and not as formal statistical superiority.

Figure 3(a) separates RMSE from MAE across the seven models and shows a monotonic improvement as increasingly explicit temporal structure is introduced. STID-Mixer produced an RMSE of 0.118 g/10 min and an MAE of 0.083 g/10 min, compared with 0.151 and 0.107 g/10 min for CatBoost and 0.129 and 0.091 g/10 min for the temporal convolutional network. The 21.9% RMSE reduction relative to CatBoost is larger than the corresponding MAE reduction, indicating that residence-aware mixing primarily shortened the high-error tail rather than shifting every prediction by the same amount. Figure 3(b) gives the complementary fit and tolerance view. R^2 increased from 0.895 for CatBoost to 0.936 for STID-Mixer, while the within-tolerance rate rose from 85.8% to 91.7%. A six-point tolerance gain is approximately equivalent to an increase of about 217 extra test windows in the engineering band, and thus has an operational interpretation other than an increase in the squared-error statistic. The model ranks remained the same after calculating the errors of the low-, medium- and high-index groups separately. The advantage was smallest in the central area because all nonlinear models have rich examples there, and largest near the grade transition, where the temporal position is still present in the aggregated features that have been lost due to aggregation. RMSE and MAE can also be used to diagnose, and if they are far apart, then there are some serious errors. Its reduction for STID-Mixer indicates that the model has improved the general window and, at the same time, also reduced the number of high-quality warning cases or retrospective process investigation cases.

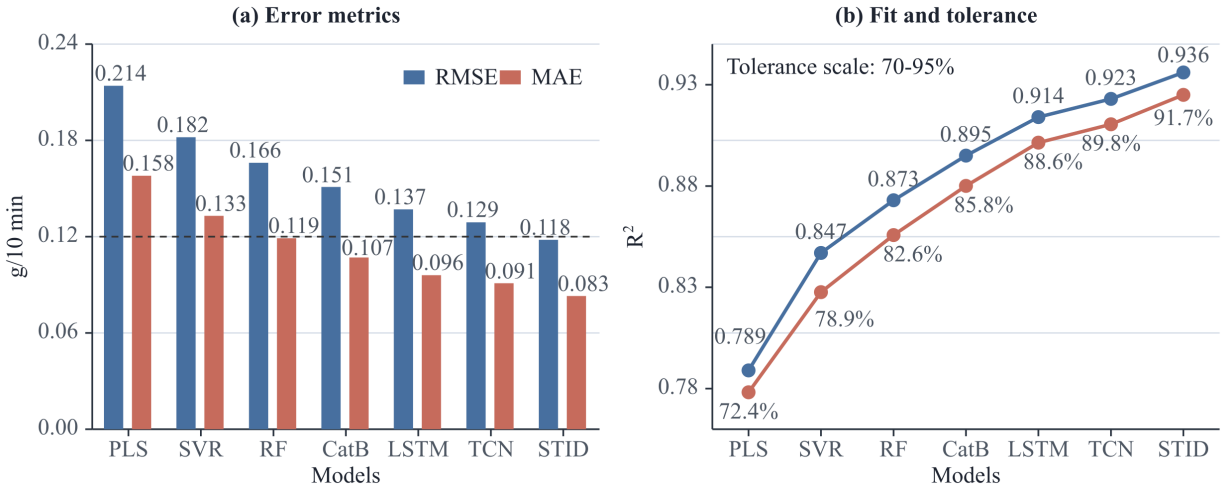


Figure 3. Benchmark accuracy comparison: (a) RMSE and MAE across seven models; (b) R^2 and within-tolerance ratio on the test set

Table 1 consolidates the benchmark and exposes the accuracy-latency trade-off. Partial least squares required only 0.7 ms per window but left an RMSE of 0.214 g/10 min, which is 81.4% above that of STID-Mixer. LSTM and the temporal convolutional network narrowed the error but required 11.6 and 8.9 ms per window. STID-Mixer used 0.41 million trainable parameters and completed inference in 3.4 ms, less than one third of the LSTM latency. The comparison with CatBoost is especially informative: both models capture nonlinear sensor coupling, but only STID-Mixer preserves the order of states inside the residence window. A die-pressure increase near the estimated residence center can therefore be distinguished from a similar increase at the edge of the window. This temporal distinction is consistent with the smaller residual tail observed during speed changes and explains why the gain is not restricted to a single aggregate metric. Runtime variability was also limited: the 95th-percentile STID-Mixer latency was 4.1 ms, compared with 13.8 ms for LSTM and 10.4 ms for the temporal convolutional network. This margin leaves ample time within a one-minute historian cycle for input validation, prediction logging, tolerance checking, and interface refresh. Parameter count alone does not guarantee stable deployment, but the compact matrix-based operations reduce dependence on sequential execution and simplify export to a CPU service. The benchmark therefore supports the selected architecture on both statistical and implementation grounds without treating speed as a substitute for predictive validity.

Table 1. Prediction accuracy and inference latency of baseline models and STID-Mixer

Model	RMSE	MAE	R^2	Within +/-0.20	Latency ms/window
Partial least squares	0.214	0.158	0.789	72.4%	0.7
Support vector regression	0.182	0.133	0.847	78.9%	4.8
Random forest	0.166	0.119	0.873	82.6%	7.3
CatBoost	0.151	0.107	0.895	85.8%	5.1
LSTM	0.137	0.096	0.914	88.6%	11.6
Temporal convolutional network	0.129	0.091	0.923	89.8%	8.9
STID-Mixer	0.118	0.083	0.936	91.7%	3.4

5.2. Residual Structure and Component Sensitivity

Residual structure was examined to determine whether the benchmark gain was broadly distributed. Figure 4(a) plots measured against predicted melt index for the complete test period; the fitted slope was 0.97, and the point cloud remained close to the 1:1 line across most of the 0.72-4.86 g/10 min range. Dispersion widened above 4.2 g/10 min, showing that the model slightly compressed the upper tail. Figure 4(b) resolves absolute error by grade. Median error remained below 0.09 g/10 min for G1-G5 but reached 0.118 g/10 min for G6, whose higher torque baseline and narrower thermal operating window amplify small alignment errors. Figure 4(c) shows a residual density centered near zero, with 94.1% of residuals between -0.20 and 0.20 g/10 min; the remaining mass is mildly right-skewed rather than bimodal, so the principal issue is occasional underprediction rather than a missing grade cluster. Figure 4(d) separates operating regimes: RMSE was 0.109 in steady operation, 0.143 during transitions, 0.136 under pressure drift, and 0.116 at high throughput. Prior extrusion monitoring studies identify transport uncertainty during transitions as a distinct source of

residual broadening [20]. The scatter is densest between 1.4 and 3.3 g/10 min because four grades occupy that interval, yet the residual width did not collapse there enough to explain the overall score by class frequency alone. At the low end, a few overpredictions occurred immediately after high-viscosity production, consistent with lingering material memory. At the high end, underprediction was associated more often with fast throughput changes than with stationary thermal offsets. These observations link the visible tails to identifiable operating events and provide a stronger diagnostic than a single residual standard deviation.

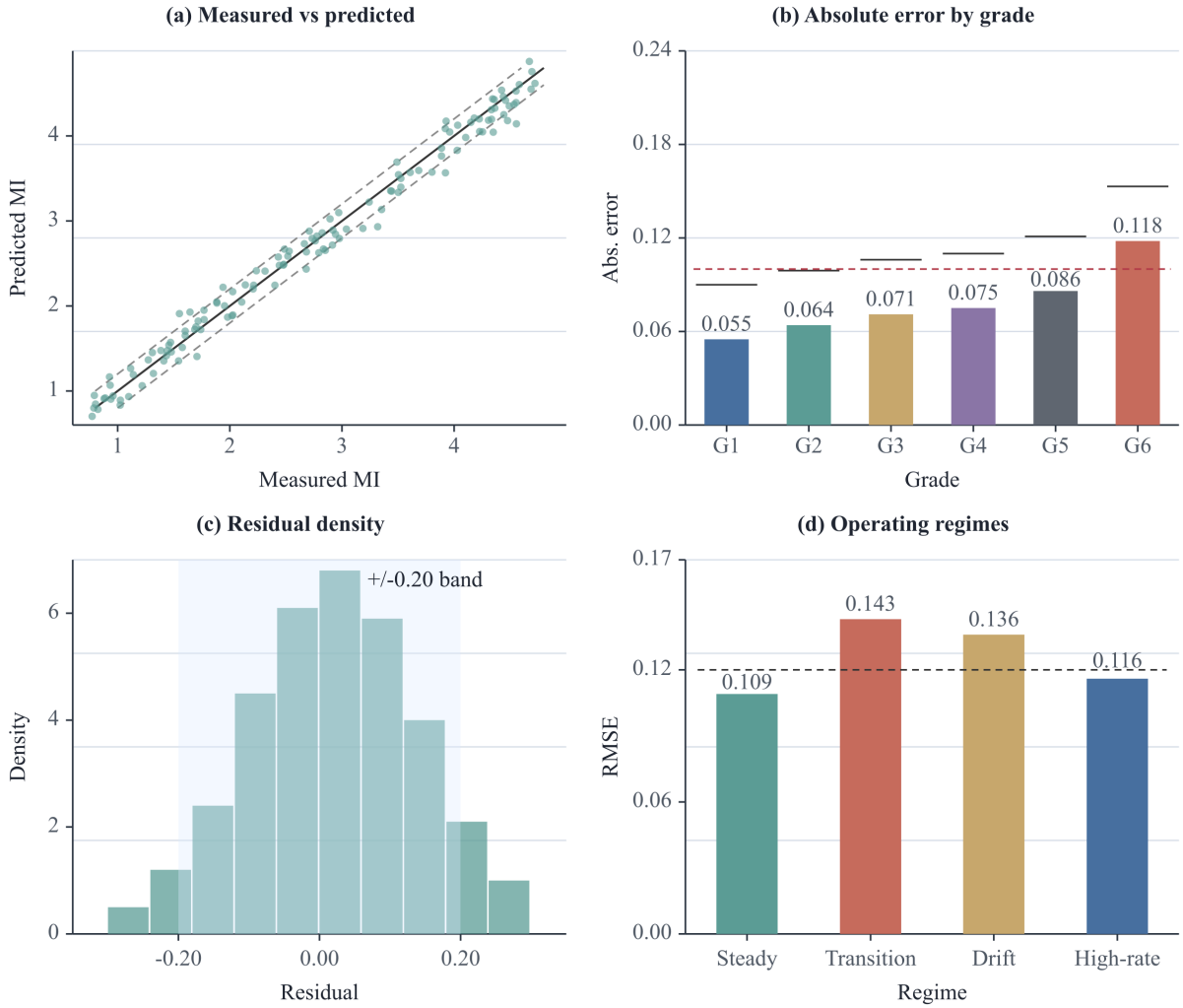


Figure 4. Error distribution and operating-regime behavior: (a) measured versus predicted melt index; (b) absolute error by product grade; (c) residual density and tolerance band; (d) RMSE under steady, transition, drift, and high-throughput regimes

Grade and operating-regime slices exhibited different failure patterns. For G6, the upper quartile of absolute error reached 0.18 g/10 min even though the median remained suitable for advisory monitoring. Accuracy decreased when thermal state and feeder stability changed simultaneously, suggesting that the high-viscosity operating region was more sensitive to small alignment errors. During screw-speed changes, pressure and torque had not fully adjusted to the new material condition, and the residual distribution shifted temporarily. Progressive pressure drift instead produced a gradual change in mean error accompanied by a modest increase in variance. These patterns are consistent with, but do not independently prove, short-lived label-alignment uncertainty and longer-lived sensor or process drift. An implementation could therefore apply different persistence rules: transition excursions may be annotated over the estimated residence interval, whereas monotonic multiwindow drift should trigger sensor-health review. Grade-specific limits may reduce nuisance alarms for G6, but additional assays near transition centers are required before such limits are adopted operationally.

Ablation was organized as a component-by-window matrix because dynamic soft sensors depend jointly on representation and memory length [21]. Figure 5(a) shows that the complete configuration reached its minimum validation RMSE of 0.118 g/10 min at 30 minutes. Removing grade normalization raised the best comparable error to 0.143 g/10 min, confirming that pooled absolute recipes obscure within-grade quality movement. Excluding residence weighting raised error to 0.136 g/10 min, with the largest loss after feeder-rate changes. Sensor mixing and temporal mixing contributed differently: without sensor mixing, the model underreacted to pressure-to-throughput variation;

without temporal mixing, it responded to current pressure and torque but missed slower thermal memory. A 15-minute window omitted part of the delayed thermal response, whereas 45- and 60-minute windows admitted stale states from earlier local conditions. The optimum is therefore tied to the material transport range of this extruder and should be re-estimated rather than copied unchanged to equipment with another screw geometry or throughput envelope. The matrix also reveals interaction between preprocessing choices. Grade normalization alone removed a broad offset but could not recover delayed feeder effects; residence weighting without normalization still mixed incompatible recipe levels. Once both were present, sensor mixing contributed more at 30 minutes than at 15 minutes because it could compare pressure, torque, and throughput over a complete material path. Temporal mixing supplied the final reduction by assigning different influence to early and late positions. This pattern supports the decomposition used in STID-Mixer: the blocks are complementary operations on the same process history rather than redundant sources of model capacity.

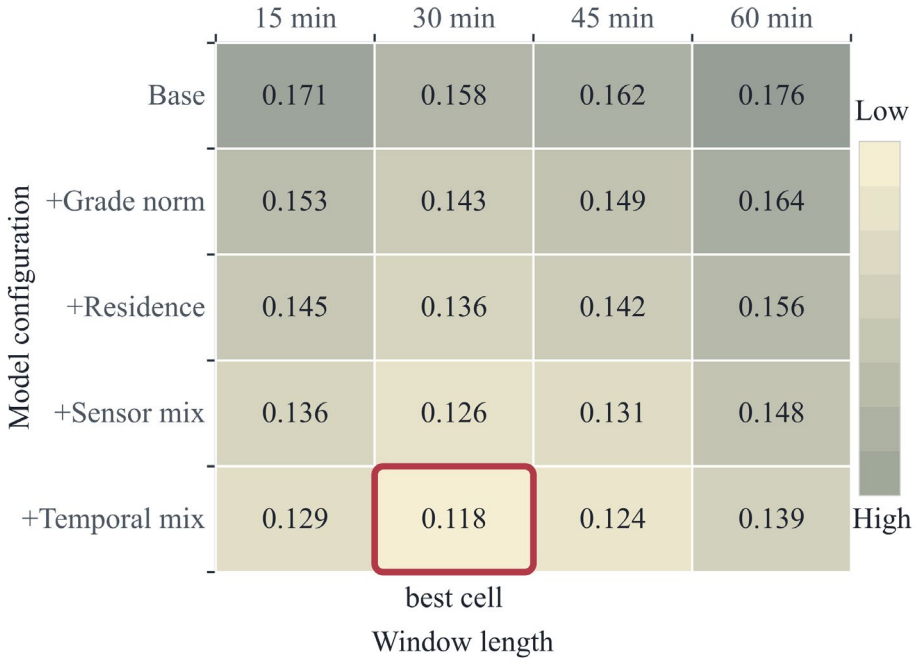


Figure 5. Ablation analysis: (a) validation RMSE under grade normalization, residence weighting, sensor mixing, and temporal mixing across candidate window lengths

Permutation and drift tests supplied process-level checks on what the network used. Permutation importance is a standard post-training diagnostic for estimating how strongly prediction error changes when one input group is disrupted [22]. Figure 6(a) identifies die pressure, screw torque, melt temperature, pressure-to-throughput ratio, and feeder stability as the leading groups; their permutations increased RMSE by 0.041, 0.035, 0.031, 0.027, and 0.024 g/10 min, respectively. The ordering is physically coherent because these channels describe resistance to flow, mechanical work, thermal state, and material delivery. Figure 6(b) then isolates die-pressure drift. Mean residual stayed below 0.02 g/10 min through 4% drift, reached 0.032 g/10 min at 6.5%, and climbed to 0.071 g/10 min at 10%; RMSE simultaneously increased from 0.118 to 0.158 g/10 min. Sensor-health monitoring is commonly paired with industrial soft sensors because an unchanged numerical output can conceal degraded input semantics [23]. Here, 6.5% is an empirical warning point for this dataset, not a universal physical limit. Lower-ranked channels still served as context. Adapter pressure and motor current added smaller marginal increases of 0.018 and 0.014 g/10 min, while vacuum contributed 0.010 g/10 min after grade normalization. Their low permutation scores do not establish physical irrelevance; correlated channels can substitute for one another, reducing the measured marginal effect. The drift curve is therefore more actionable than a strict feature ranking. It quantifies the operating distance at which a familiar high-importance channel begins to bias the output and can be implemented as a monitored condition beside missing-value and stale-tag checks.

5.3. Robustness and Online Replay

Robustness tests were designed around two practical shortages: incomplete laboratory coverage and imperfect sensors. Label-sparsity and additive-noise protocols are widely used to distinguish data efficiency from nominal test accuracy [24]. Figure 7(a) shows that RMSE decreased from 0.146 g/10 min with 40% of training labels to 0.124 with 70% and 0.118 with the complete set. The smooth curve indicates gradual loss rather than a threshold collapse, although sparse labels cannot replace new assays after a resin, additive, or die change. Figure 7(b) adds channel-wise Gaussian noise at 0.5%, 1.0%, and 2.0% of scale. RMSE rose to 0.122, 0.126, and 0.132 g/10 min, respectively, so the largest perturbation added 0.014 g/10 min without destabilizing the predictor. Pressure and torque accounted for most of this degradation, whereas slow temperature channels were partly protected by temporal aggregation. Tolerance coverage remained above

88% at the 2.0% noise setting, although the residual tail became visibly heavier. Repeating the perturbation with one sensor group at a time showed that simultaneous pressure and torque noise caused a larger loss than the sum of their isolated effects, reflecting their coupled role in the inferred viscosity state. This interaction cautions against validating sensors independently during commissioning. A calibration check that perturbs or compares related channels together is more likely to expose a combination that moves the latent process state outside the training distribution.

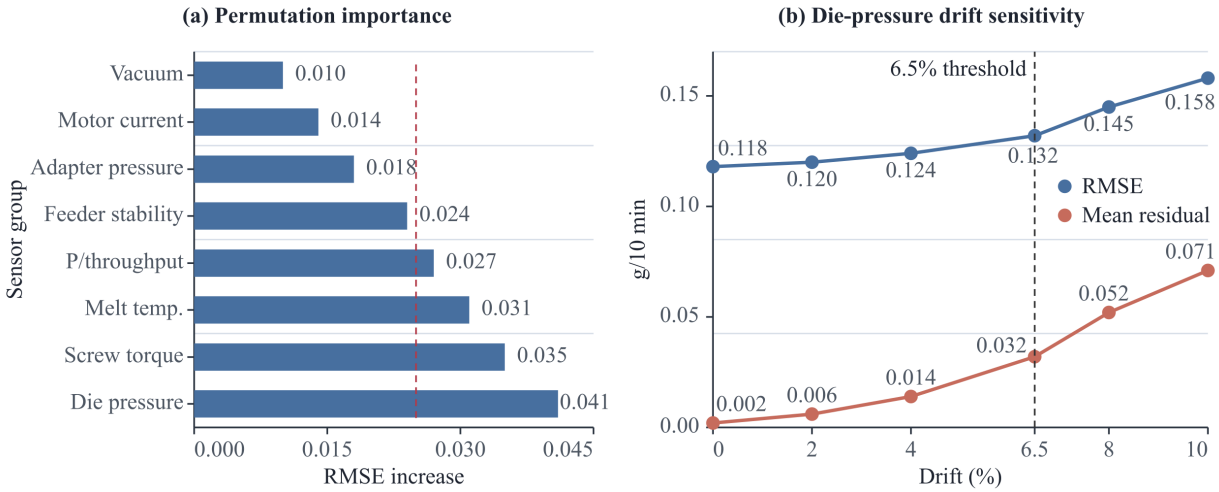


Figure 6. Sensor contribution and drift sensitivity: (a) permutation importance by sensor group; (b) residual mean and RMSE under increasing die-pressure drift

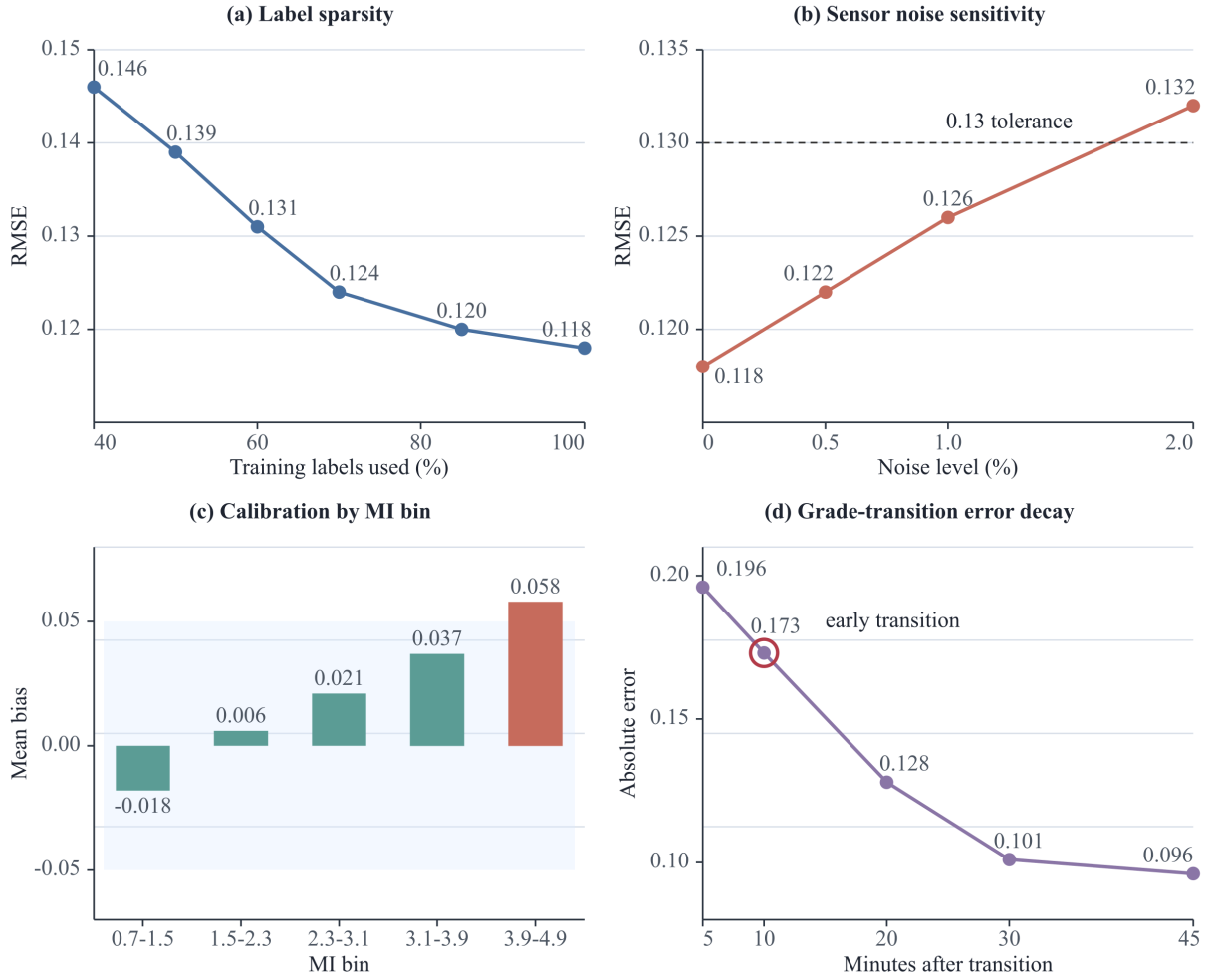


Figure 7. Robustness tests: (a) label-sparsity curve; (b) noise-level sensitivity; (c) calibration by melt-index bin; (d) grade-transition error decay

Calibration and transition timing reveal where the robustness margin narrows. Figure 7(c) groups predictions by measured melt-index interval. Mean bias remained between -0.018 and 0.037 g/10 min for the first four bins but

increased to 0.058 g/10 min in the 3.9-4.9 g/10 min bin, matching the upper-tail compression in the scatter plot. The model is therefore well centered over the dominant range but should expose a range flag near the upper boundary. Figure 7(d) follows transition error over time. Absolute error was 0.196 g/10 min at five minutes and 0.173 at ten minutes, then declined to 0.128 at twenty minutes and 0.101 at thirty minutes. This decay agrees with a mixed-material interpretation: the sensor stream begins to describe the new recipe before the sampled polymer has completely traversed the line. A transition-aware display can retain the estimate while marking its temporarily wider uncertainty instead of suppressing all output during grade changes. The bias in the highest index bin involved fewer samples and wider confidence bounds, so it should be treated as an identified calibration weakness rather than as evidence of a fixed offset correction. Applying a global correction would damage the well-centered middle bins. A better response is to collect targeted assays at the boundary and update the grade-specific representation. During transitions, the error curve suggests a separate temporal rule: predictions can remain visible from the first minute, but the interface can display a declining transition-risk indicator until pressure, speed, and feeder stability have remained inside the new grade bands for the required residence period.

Perera et al. [25] showed that online polymer-extrusion soft sensors must balance rapid response with sufficiently smooth estimates for control use. Figure 8(a) places the evaluated methods in latency-error space. STID-Mixer required 3.4 ms per window and achieved an RMSE of 0.118 g/10 min; the 5 ms boundary shown in the figure is a project-side interface target rather than a universal industrial standard. CatBoost required 5.1 ms with an RMSE of 0.151 g/10 min, whereas LSTM and the temporal convolutional network occupied higher-latency regions. Figure 8(b) presents a 180-minute replay containing six laboratory checkpoints. The mean absolute checkpoint deviation was 0.086 g/10 min, and two intervals near a feeder-rate disturbance briefly exceeded the ±0.20 g/10 min warning band. The prediction changed smoothly rather than through isolated one-minute spikes, but this retrospective replay does not establish prospective control performance. Hardware and software identifiers should be reported with future latency measurements so that deployment claims can be reproduced across computing environments.

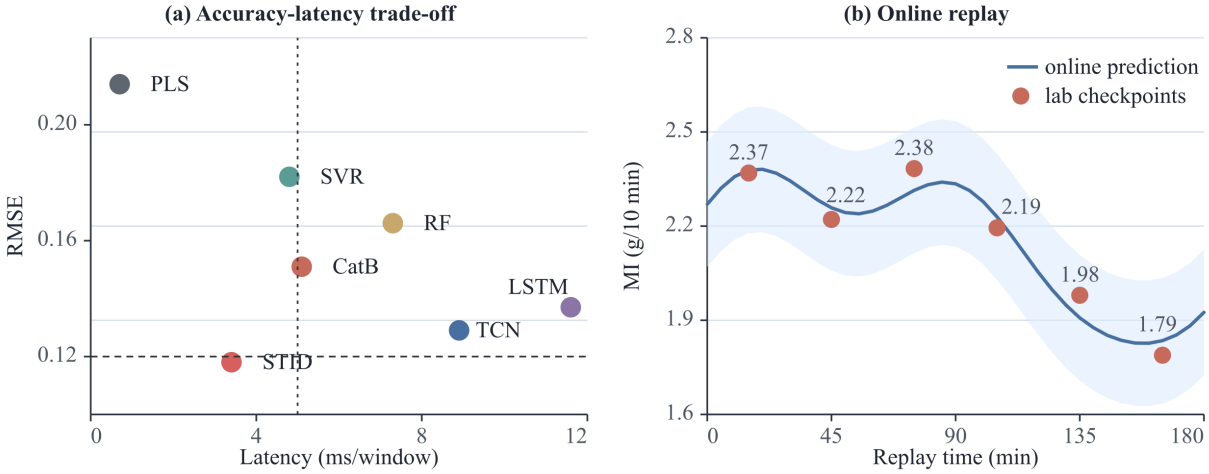


Figure 8. Online suitability analysis: (a) latency versus RMSE trade-off for all models; (b) 180-minute online replay with laboratory checkpoints and warning bands

Deployment is still subject to the integrity of the data path. Long-term soft-sensor maintenance guidance focuses on timestamp synchronization, stable tag semantics, unit control, and explicit handling of sensor replacement [26]. Therefore, the online pipeline should reject windows with extended missing tags, impossible ranges, stale values, unknown grade codes or sudden scaling changes prior to inference. A small gap will only use the last valid standardized value from the documented hold period. Several windows will be opened before a quality hold is triggered, and otherwise, a single excursion due to sampling errors or temporary noise will occur. Given the occurrence of a pressure-drift condition, an empirical 6.5% threshold will be employed by the monitoring layer in conjunction with feeder stability and transition status to distinguish between maintenance drift and a planned recipe modification. Laboratory verification has not been carried out for a new resin batch, additive package, screw condition or die configuration. Save the numerical estimate along with the model version, grade code, window validity, transition status, drift flags and timestamps of the source tags. The above attributes can be used for back-comparison with the stored data to prevent the apparent accuracy of the value from being out of context. After modifying or extending a tag, a shadow mode can be used to compare the new stream with the old baseline and display the prediction to the user. The replay results can be used to display the tested range continuously, but they do not authorize autonomous recipe correction or release decisions without the construction of a laboratory and quality procedures for the plant.

6. Conclusion

This study developed and evaluated STID-Mixer for melt-index prediction from extrusion sensor streams, demonstrating that residence-aware sensor-time decomposition can combine accurate quality estimation with practical online inference.

1. On 18,420 one-minute windows covering 17 raw sensors, 11 derived descriptors, and six polymer grades, STID-Mixer achieved an RMSE of 0.118 g/10 min, an MAE of 0.083 g/10 min, and an R^2 of 0.936. A total of 91.7% of test predictions remained within the +/-0.20 g/10 min engineering band, indicating that the model maintained both low average error and strong tolerance coverage across the chronological test period.

2. The performance gain was associated with complementary process and model effects rather than additional network depth. Relative to CatBoost, RMSE decreased by 21.9%, while inference latency remained at 3.4 ms per window compared with 11.6 ms for the LSTM baseline. Grade normalization removed recipe-level offsets, residence weighting improved delayed sample alignment, and the sensor and temporal mixers reduced errors arising from coupled variables and process memory. Prediction error was lowest during steady operation and increased during early grade transitions and progressive die-pressure drift.

3. The principal methodological contribution is the separation of cross-sensor interaction learning from residence-aware temporal mixing within a compact soft-sensing architecture. The evidence is limited to one extrusion line, six grades, and the operating ranges represented in the historian. Errors were larger for grade G6 and during the initial stage of product transitions, indicating a need for denser assay alignment and targeted sampling in these regions. The 6.5% pressure-drift warning point is an empirical stress-test result for this dataset rather than a general process limit and must be recalibrated before transfer to another line or sensor configuration.

The proposed framework provides a practical basis for continuous extrusion-quality surveillance, while cross-line validation, uncertainty-aware alarms, automated tag checking, and adaptive drift correction remain important directions for future development.

7. Declarations

7.1 Author Contributions

Hrvoja Rajšić contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Kraljevka Novčićcontributes to validation, analysis, investigation, data collection, draft preparation. Ignjata Milutinović contributes to conceptualization, validation, analysis. All authors have read and agreed with the manuscript before its submission and publication.

7.2 Funding and Acknowledgments

This research received no specific financial support from any funding agency.

7.3 Institutional Review Board Statement

Not applicable.

7.4 Informed Consent Statement

Not applicable.

7.5 Data Availability

Not applicable.

7.6 Declaration of Competing Interest

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

7.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

8. References

[1] Sheng, W., Qian, J., Song, Z., & Zhang, X. (2024). A review of just-in-time learning-based soft sensor in industrial process. The Canadian Journal of Chemical Engineering, 102(5), 1884–1898. https://doi.org/10.1002/cjce.25169

[2] Offermans, T., Szymańska, E., Souza, F. A., & Jansen, J. J. (2024). Process expert knowledge is essential in creating value from data-driven industrial soft sensors. Computers & Chemical Engineering, 183, 108602. https://doi.org/10.1016/j.compchemeng.2024.108602

[3] Hütter, S., Simonin, Y., Mook, G., & Halle, T. (2024). Soft sensor system for in-process eddy current microstructure characterization. Journal of Sensors and Sensor Systems, 13(2), 237–244. https://doi.org/10.5194/jsss-13-237-2024

[4] Zhang, Z., Yang, Y., Hou, Y., Ou, D., & Xu, L. (2024). Soft sensor for melting flow rate prediction based on data-enhanced classification method. Measurement Science and Technology, 35(12), 125122. https://doi.org/10.1088/1361-6501/ad7c6e

[5] Kay, H., Kay, S., Mowbray, M., Lane, A., Mendoza, C., Martin, P., & Zhang, D. (2024). Constructing a symbolic regression-based interpretable soft sensor for industrial data analytics and product quality control. Industrial & Engineering Chemistry Research, 63(9), 4083–4092. https://doi.org/10.1021/acs.iecr.3c04021

[6] Wei, B., Tan, S., Zhang, Q., & Zhou, H. (2024). A hybrid soft sensor for key product yield of FCC unit based on deep learning framework driven by data and process mechanism. Chemical Engineering Research and Design, 202, 429–443. https://doi.org/10.1016/j.cherd.2024.01.015

[7] Gao, S., Xu, J., Ma, Z., Tian, R., Dang, X., & Dong, X. (2024). Research on modeling of industrial soft sensor based on ensemble learning. IEEE Sensors Journal, 24(9), 14380–14391. https://doi.org/10.1109/jsen.2024.3375072

[8] Ni, M., & Li, S. (2024). A data-driven LSTMSCBLS model for soft sensor of industrial process. Measurement Science and Technology, 35(10), 106201. https://doi.org/10.1088/1361-6501/ad5ab8

[9] He, Y. L., Lv, S. H., Zhu, Q. X., & Lu, S. (2024). Novel multiattribute space-based LSTM for industrial soft sensor applications. IEEE Transactions on Industrial Informatics, 20(3), 4745–4752. https://doi.org/10.1109/tii.2023.3316289

[10] Tuo, B., Zhao, X., Sun, K., Liu, K., & Hui, Y. (2024). Soft sensor model for nonlinear dynamic industrial process based on GraphSAGE-IMATCN. Process Safety and Environmental Protection, 191, 1131–1147. https://doi.org/10.1016/j.psep.2024.08.023

[11] Wang, X., Qi, X., & Zhang, Y. (2024). Deep learning with local spatiotemporal structure preserving for soft sensor development of complex industrial processes. Applied Soft Computing, 164, 111974. https://doi.org/10.1016/j.asoc.2024.111974

[12] Ma, L., Wang, M., & Peng, K. (2024). A spatiotemporal industrial soft sensor modeling scheme for quality prediction with missing data. IEEE Transactions on Instrumentation and Measurement, 73, 1–10. https://doi.org/10.1109/tim.2024.3400358

[13] Fang, Z., Gao, S., Dang, X., Dong, X., & Wang, Q. (2024). Transformer enhanced by local perception self-attention for dynamic soft sensor modeling of industrial processes. Measurement Science and Technology, 35(5), 055123. https://doi.org/10.1088/1361-6501/ad25dd

[14] Zhang, L., Ren, G., Li, S., Du, J., Xu, D., & Li, Y. (2025). A novel soft sensor approach for industrial quality prediction based TCN with spatial and temporal attention. Chemometrics and Intelligent Laboratory Systems, 257, 105272. https://doi.org/10.1016/j.chemolab.2024.105272

[15] Liu, Q., Jia, M., Guo, X., Liu, Y., Gao, Z., & Xu, L. (2024). Two-dimensional LSTM soft sensor using noisy process data. Measurement Science and Technology, 35(8), 085001. https://doi.org/10.1088/1361-6501/ad4384

[16] Wu, J., Shen, F., & Ye, L. (2024). Data augmentation using time conditional variational autoencoder for soft sensor of industrial processes with limited data. IEEE Transactions on Instrumentation and Measurement, 73, 1–14. https://doi.org/10.1109/tim.2024.3427765

[17] Kay, S., Kay, H., Mowbray, M., Lane, A., Mendoza, C., Martin, P., & Zhang, D. (2024). Integrating transfer learning within data-driven soft sensor design to accelerate product quality control. Digital Chemical Engineering, 10, 100142. https://doi.org/10.1016/j.dche.2024.100142

[18] Xie, S., Yu, Y., Xie, Y., & Tang, Z. (2024). Sensitive feature selection for industrial flotation process soft sensor based on multiswarm PSO with collaborative search. IEEE Sensors Journal, 24(10), 17159–17168. https://doi.org/10.1109/jsen.2024.3381837

[19] Zhai, R., Zheng, J., Song, Z., & Ge, Z. (2024). Reliable soft sensors with an inherent process graph constraint. IEEE Transactions on Industrial Informatics, 20(6), 8798–8806. https://doi.org/10.1109/tii.2024.3372013

[20] Cao, L., Wang, J., Su, J., Luo, Y., Cao, Y., Braatz, R. D., & Gopaluni, B. (2025). Comprehensive analysis on machine learning approaches for interpretable and stable soft sensors. IEEE Transactions on Instrumentation and Measurement, 74, 1–17.

https://doi.org/10.1109/tim.2025.3556830

[21] Cao, L., Su, J., Conde, E., Siang, L. C., Cao, Y., & Gopaluni, B. (2025). A novel automated soft sensor design tool for industrial applications based on machine learning. Control Engineering Practice, 160, 106322. https://doi.org/10.1016/j.conengprac.2025.106322

[22] Zhang, J., Ai, M., Tang, Z., Xie, Y., Ma, J., & Wu, J. (2025). A sentinel-based adaptive hybrid soft sensing method for industrial process monitoring. IEEE Transactions on Instrumentation and Measurement, 74, 1–13. https://doi.org/10.1109/tim.2025.3554879

[23] Li, W., Dai, W., Nan, J., & Li, B. (2025). Semi-supervised incremental soft sensor model with spatiotemporal graph regularization for process industry. IEEE Transactions on Instrumentation and Measurement, 74, 1–15. https://doi.org/10.1109/tim.2024.3500040

[24] Jiang, D., Yang, H., Cao, H., & Xu, D. (2025). A missing data imputation method for industrial soft sensor modeling. Journal of Process Control, 152, 103485. https://doi.org/10.1016/j.jprocont.2025.103485

[25] Perera, Y. S., Li, J., & Abeykoon, C. (2025). Melt viscosity control in polymer extrusion using nonlinear model predictive control with neural state space modelling and soft sensor feedback. Journal of Process Control, 155, 103556. https://doi.org/10.1016/j.jprocont.2025.103556

[26] Hartner, R., Kozek, M., & Jakubek, S. (2026). Multi-task learning with state propagation for quality forecasts in polymer extrusion lines. Journal of Intelligent Manufacturing, 37(4), 1701–1715. https://doi.org/10.1007/s10845-025-02616-2