Journal of Applied Automation Technologies

Research Paper

Mamba-TSMixer Model for Tool Chatter Prediction Based on Spindle Acoustic Emission Signal Fusion

Karolina Tatiana Kuśowa[1] and Ludwik Lupa[1,*]

[1] Faculty of Mechatronics and Automation, Cracow University of Technology, Krakow, 31-155, Poland
*Corresponding author: ludwik.l@pk.edu.pl

Abstract. Reliable warning of regenerative chatter is still an issue because spindle speed, engagement and tool wear change the observable vibration signature. Develop a causal Mamba-TSMixer model in this paper that integrates spindle-mounted acoustic-emission data with rotational and process context for horizon-based chatter prediction. Raw emission signals are converted into synchronous envelope, impulsiveness, spectral and order-aware descriptors. A selective state-space path retains the extended-time dependency; a temporal-channel mixer handles local cross-feature correlations; and cross-gated fusion avoids context-inconsistent emission bursts. There are 1,260 milling experiments, five speed ranges, four wear conditions and two unseen-range experiments. At a prediction horizon of 200ms, the model has an F1 score of 0.943 for events, an area under the receiver operating characteristic curve (AUC) of 0.961, and a median warning lead of 184ms. Compared with the best-performing baseline, event F1 is up by 3.8% and false alarms have dropped from 0.31 to 0.18 per cutting minute. Performance is still 0.912 F1 after masking one group of acoustic-emission features, and causal inference needs 1.8ms per update. Ablation shows that selective memory and gated fusion contribute to the gain independently. Run the grouped split and retention speed-wear combinations to further test whether the predictor has exceeded the familiar cutting adjacent window. Based on the above experiments, high-frequency spindle emission can be used as an early warning indicator under different machining conditions, and models have been developed for both long-term and short-term analysis. The constructed structure offers a high-performance platform for the unstable operating environment of real-time monitoring.

Keywords: *Time-Series Modeling, Mamba Network, TSMixer, Acoustic Emission, Predictive Manufacturing*

Received on 18 February 2024, Accepted on 19 October 2024, Published on 28 October 2024

Introduction

Chatter is a self-excited instability that reduces surface quality, increases tool damage, and limits the selection of good-milling parameters. Classical stability analysis explains the regenerative origin [1], but shop-floor prediction is difficult due to structural dynamic drift after tool change and thermal load. Spindle-integrated sensing provides a small-scale observation point [2], and acoustic emission (AE) can detect high-frequency elastic activity that often changes before it appears in low-frequency vibration [3]. AE monitoring has therefore been applied to tool-condition assessment [4], transient fracture detection [5], and machining-state recognition [6]. The above achievements are not directly linked to prediction; rather, the consequences of engagement, chip segmentation, coolant events and bearing noise can all be energetic bursts similar to those preceding an unstable fracture.

Data-driven chatter monitors have gone from handcrafted spectra to convolutional and recurrent sequence learners. Time-frequency representations improve separation of cutting states [7], and deep convolutional features reduce dependence on manually selected bands [8]. Recurrent predictors maintain the temporal order [9], attention mechanisms extend the effective observation window [10], and multimodal fusion improves robustness against ambiguity in a single channel [11]. Quadratic attention cost and recurrent execution are not suitable for dense AE updates, especially when the model needs to issue calibrated probabilities in a controller cycle [12]. A practical predictor should have retained long-term pre-onset evidence, resolved short impulsive changes, distinguished genuine dynamic growth from speed-conditioned emission, and maintained causal latency.

This work addresses this problem by proposing a dual-path Mamba-TSMixer architecture for synchronous spindle AE and process context. The three contributions are as follows: a leakage-resistant horizon labeling and signal construction procedure; a causal fusion network that combines selective state-space memory with temporal and channel mixing; and an evaluation that treats early-warning lead time, event-level accuracy, calibration, robustness and runtime as joint engineering requirements. The study is organized around a 200-millisecond prediction time frame and deliberately does not use random partitions of adjacent windows; therefore, the reported performance reflects the transfer under conditions not seen during the fitting process.

Related Work

Chatter Sensing and Acoustic-Emission Monitoring

The stable blade theory remains the reference system for analyzing regenerative chatter [13]. Operational monitoring is still based on measured responses because modal parameters and cutting coefficients change during service. Accelerometers are directly structural measurements [14]; microphones are used for non-contact collection [15]; and spindle-current signals can be acquired with little hardware modification [16]. AE sensors expand the observable bandwidth and are sensitive to tool-workpiece contact phenomena [17]. Sensitivity is also a weakness; resonance amplification, preload and lubrication change amplitude distribution, and impulsive disturbances occur [18]. Therefore, robust AE usage depends on normalization and contextual conditions, rather than a universal energy threshold.

These sensing routes observe different places along the same instability chain from an engineering perspective. Structural vibration is closer to the final dynamic response, and acoustic emission is closer to intermittent contact, friction and micro-fracture at the cutting zone. The latter can thus provide earlier evidence, but its physical transfer path through the tool, spindle, bearings and housing is more susceptible to installation. A good predictor should not assign a large emission amplitude to chatter as evidence for this. Consideration should be given to durability, frequency organization, consistency with the tooth transmission structure, and consistency with the current engagement state. In the evaluation, separate the impact of the inlet, the transient at the outlet, the broadband growth related to wear, and the actual regeneration development. Therefore, synchronous context variables, robust standardization, quality masks, and event-level scoring have been introduced. It can also explain why this study intends to predict the start of a problem rather than discovering it after the fact.

Efficient Temporal Learning and Multisensor Fusion

One-dimensional convolutional networks have been used for machinery sequences due to their computationally simple parallel receptive fields [19]. The gated recurrent unit model has temporal dependencies and maintains sequential state updates [20]. Transformers have improved long-range interaction modeling [21], but attention memory increases with the sequence length. Structured State-Space Models offer another path for long-context representation [22], and selective state updates enable input-dependent information retention [23]. Mixer architectures separately transform the temporal and feature axes with simple Multi-Layer Perceptrons [24]. Most existing chatter studies do not combine selective long-memory dynamics with explicit local channel mixing, and few have tested the utility of this fusion under missing AE features and unseen machining regimes. Another limitation is that many reported fusion pipelines connect sensor features before temporal modeling; therefore, the network cannot clearly determine the timing of conflicts between high-frequency emissions and slower operational contexts. Separate local interaction modelling from long-term memory dynamics to find a more causal, quality-aware fusion path.

The last Design problem is not to select the model with the longest theoretical basis. Chatter precursors are slow changes over several spindle revolutions with short impulsive structures within individual updates. A single temporal operator may maintain one scale but smooth the other. Early feature concatenation also assumes that all channels are equally reliable; however, under conditions such as packet loss, clipping, tachometer uncertainty or sudden process changes, this assumption is not met. Therefore, the architecture in this paper has divided the

https://doi.org/10.64972/jaat.2024v2.341p33e:456-469

functions of selective memory, local temporal-channel interaction and quality-aware fusion. This division supports causal state caching, provides a direct sensor-ablation path, and makes the computational path more amenable to audit.

Signal Construction and Fusion Protocol

Synchronized Acquisition and Horizon Labels

Samples at 1 MHz are taken of the acquisition chain for spindles and controller variables at 1 kHz. Each prediction update is a causal context of 512ms and advances by 64ms. Tachometer edges align AE blocks with spindle revolutions; commanded feed, spindle speed, radial immersion, axial depth, and cumulative cutting time constitute the context stream. Chatter onset is identified by the agreement of order-synchronous vibration growth and persistent sideband energy, then moved back to the forecast horizon. The binary target for update index t is defined as

$$y_t^{(h)} = \begin{cases} 1, & 0 < \tau_{\mathrm{on}} - t\Delta \leq h, \\ 0, & \text{otherwise}. \end{cases} \quad \text{Eq.(1)}$$

where Δ is the 64ms update interval, τ_{on} is the annotated onset, and h is 200ms. Discard Windows that cross run boundaries. Group the training, validation and test sets by the complete cutting run to avoid splitting adjacent waveform blocks.

$$\tilde{x}_{t,k} = \frac{x_{t,k} - \mathrm{median}_k}{1.4826\mathrm{MAD}_k + \varepsilon} \quad \text{Eq.(2)}$$

Only fit robust scaling on the training run. It reduces the effect of isolated AE bursts and keeps the continuous distribution change. Speed is a quantity and its rate of change; angular position converts the tooth-passing frequency into an explicit context variable.

Label construction also excludes the first two updates after tool engagement and the post-onset interval used to confirm the reference event. The guard region prevents entry shocks and already-developed chatter from being easy positive examples. Timestamp drift is checked once per run by matching tachometer-derived rotation with controller speed; runs exceeding one update of cumulative mismatch are reacquired rather than numerically shifted after inspection. The above controls retain the desired forecasting interpretation for each positive window.

Acoustic-Emission Descriptor Stream

There are eight overlapping microframes in each AE block. The analytic envelope maintains the transient energy growth, and crest factor, kurtosis, spectral centroid, band-energy ratio, spectral entropy, and tooth-order sideband index all indicate different alterations. Root-mean-square envelope energy is

$$E_t = \sqrt{\frac{1}{N}\sum_{n=1}^{N} a_t[n]^2} \quad \text{Eq.(3)}$$

where $a_t[n]$ is the high-pass-filtered AE sample. A normalized band contribution avoids treating a speed-driven broadband amplitude increase as chatter by itself.

$$\rho_{t,b} = \frac{\sum_{f \in B_0} |A_t(f)|^2}{\sum_f |A_t(f)|^2 + \varepsilon} \quad \text{Eq.(4)}$$

The order-aware sideband indicator measures symmetric energy around the dominant tooth-passing harmonic.

$$S_t = \frac{P(f_c - f_s) + P(f_c + f_s)}{2P(f_c) + \varepsilon} \quad \text{Eq.(5)}$$

All descriptors are kept as a short-time grid and not reduced to a single statistic. Therefore, it will not be mistaken for a single impulse and thus won't expand gradually over several microframes.

Descriptor extraction is implemented as a streaming operation. The anti-aliasing and high-pass stages maintain the filter state between updates, and only the newest 64ms of samples are transformed after the first window.

Spectral estimates use a Hann taper and zero padding only for interpolation; no future samples are in the causal block. Envelope percentiles are retained with energy because two signals of the same root-mean-square level can have different burst occupancies. Normalize the spectral peaks using the Nyquist frequency and reference all order-related features to the measured tachometer speed, rather than the commanded speed. Thus, a momentary spindle-speed tracking error will not be identified as a change in chatter frequency.

A Quality Mask Accompanying All Descriptors. Clipping is triggered when more than 0.2% of the raw samples fall within the converter rails, packet loss is detected by sequence counter reduction, and tachometer confidence drops due to consecutive pulse interval violations of the acceleration bound. Missing scalar values are replaced by the training median only after the corresponding mask bit has been set. Therefore, the model can distinguish between an ordinary central value and an imputed observation. Masking of features is used during training with a probability of 0.08 to expose the fusion layer to real sensor degradation without completely disregarding the AE information.

Causal Fusion Protocol

The AE grid and context stream are projected separately, marked with normalized time and phase, and then passed to two model paths. A learned quality vector is a clipping rate, missing samples and tachometer confidence. The whole processing sequence is shown in Figure 1; the diagram position is intentionally left empty for external artwork.

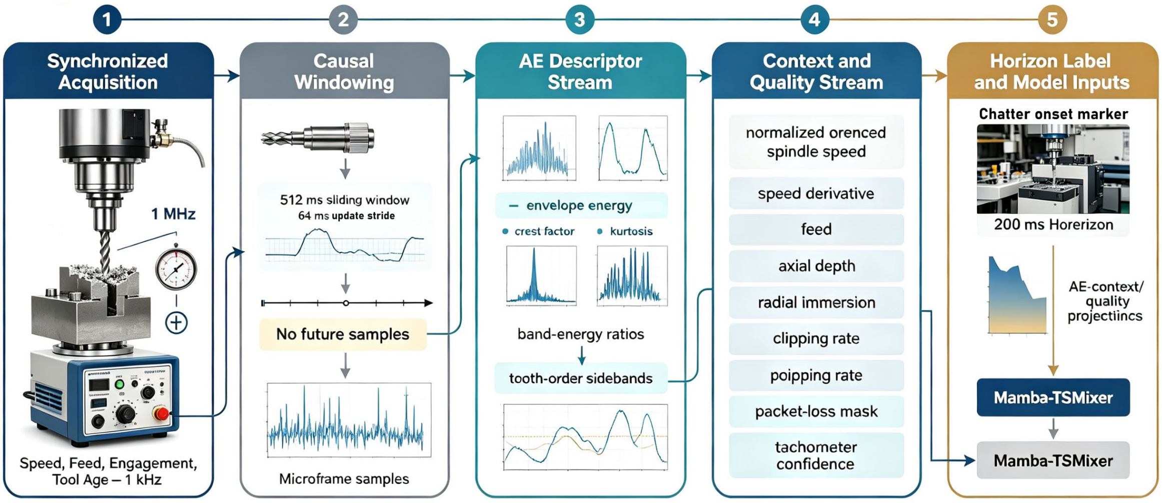

Figure 1. Causal signal-construction and synchronization workflow from spindle AE acquisition to horizon labels and model-ready streams.

$$\mathbf{z}_t^{(0)} = \mathbf{W}_a\mathbf{a}_t + \mathbf{W}_c\mathbf{c}_t + \mathbf{W}_q\mathbf{q}_t + \mathbf{b} \qquad \text{Eq.(6)}$$

Here $\mathbf{a}_t$, $\mathbf{c}_t$, and $\mathbf{q}_t$ denote AE, machining-context, and signal-quality vectors. Their separate projections make sensor ablation possible without redefining the input dimension.

The projection stage keeps the same interface when a descriptor group is not available. The missing groups are replaced by their training set centers, and the corresponding quality positions clearly indicate the replacements. Context values are only kept in the controller update interval and are never interpolated from a later command. In mini-batch construction, the full causal sequences remain ordered and state caches are initialized at the run boundary. Feature group masking is performed after normalization to prevent manual dropping from altering the statistics of the scaler. These implementation details ensure that the training behavior remains consistent with deployment and conduct robustness testing by removing information sources without changing the network dimensions. They also prevent the model from inferring missingness indirectly through extreme numerical fill values.

https://doi.org/10.64972/jaat.2024v2.341p33e:456-469

Mamba-TSMixer Prediction Architecture

Selective State-Space Path

The first path is a causal selective state-space encoder. Its transition and input parameters depend on the current fused token, enabling the model to retain weak pre-onset growth while rapidly forgetting isolated context-inconsistent impulses.

$$\mathbf{h}_t = \overline{\mathbf{A}}_t \mathbf{h}_{t-1} + \overline{\mathbf{B}}_t \mathbf{z}_t, \mathbf{u}_t = \mathbf{C}_t \mathbf{h}_t + \mathbf{D}\mathbf{z}_t \quad \text{Eq.(7)}$$

Discretized matrices are generated with a positive input-dependent step, which keeps state evolution stable across speed ramps.

$$\Delta_t = \mathrm{softplus}(\mathbf{W}_\Delta \mathbf{z}_t + \mathbf{b}_\Delta), \overline{\mathbf{A}}_t = \exp(\Delta_t \mathbf{A}) \quad \text{Eq.(8)}$$

Residual normalization and layer-wise causal convolution are performed before state updates. Four blocks of 96-dimensional hidden representations provide a receptive field of 512 ms without generating a quadratic attention matrix.

$$\mathbf{m}_t = \mathrm{SiLU}[\mathrm{Conv}_{\mathrm{causal}}\,(\mathrm{LN}(\mathbf{z}))_t] \odot \mathbf{u}_t + \mathbf{z}_t \quad \text{Eq.(9)}$$

Each block has a pre-normalized residual structure, so even if the acoustic emission energy changes dramatically, the state updates occur within a bounded feature space. The learned step size controls how much the retained history is changed by the current token; a small step size keeps accumulating evidence slowly, and a large step size adapts quickly during a real change. The state matrix is constrained to stable continuous-time dynamics before discretization, and the residual branch does not need to contain high-frequency information that is not required by the long memory. Only the newest token and cached state are used at deployment. Therefore, computation does not increase with the number of previous updates. Reset signals are associated with run boundaries, tool changes and acquisition restarts, and thus do not carry over latent evidence from the previous cut. The above situation is more pronounced in the same production plan where stable and unstable operations alternate.

Temporal-Channel Mixer and Cross-Gating

The second path performs temporal mixing of the eight most recent updates and channel mixing of descriptor groups. Temporal mixing is local evolution, while channel mixing learns interactions such as the rising envelope energy combined with increased ordered sidebands and stable participation.

$$\mathcal{T}(\mathbf{Z}) = \mathbf{Z} + [\mathbf{W}_{t2}\sigma(\mathbf{W}_{t1}\mathrm{LN}(\mathbf{Z})^\top)]^\top \quad \text{Eq.(10)}$$

A separate channel transformation follows after a paragraph-level separation, preventing two displayed equations from becoming visually crowded.

$$\mathcal{C}(\mathcal{T}) = \mathcal{T} + \mathbf{W}_{c2}\sigma[\mathbf{W}_{c1}\mathrm{LN}(\mathcal{T})] \quad \text{Eq.(11)}$$

Cross-gating uses the state-space output to assess the mixer evidence and the quality vector to reduce trust in corrupted AE descriptors.

$$\mathbf{g}_t = \sigma\big(\mathbf{W}_g[\mathbf{m}_t; \mathbf{c}_t; \mathbf{q}_t] + \mathbf{b}_g\big), \mathbf{f}_t = \mathbf{g}_t \odot \mathbf{m}_t + (1 - \mathbf{g}_t) \odot \mathcal{C}_t \quad \text{Eq.(12)}$$

Temporal mixing is carried out in the compact recent-update dimension rather than on the raw acoustic waveform, and its parameters focus on changes in descriptor shape and repetition. Channel mixing then combines physically distinct groups without assuming that adjacent feature indices have local meaning. A cross-gate is calculated independently for each update and feature channel. Signal quality can reduce the impact of noise in the acoustic evidence, and machining context may suppress energy patterns that are expected during start-up or acceleration. Residual access to the two paths prevents gate saturation by not permanently discarding one representation. Gate entropy is used to observe whether it will collapse to a single branch during training, but it is not optimized as a target. The above structure is suitable for quantifying the fusion mechanism; thus, after a clean cut or perturbation test, changes in the average gate weights can be observed and compared with prediction errors.

Forecast Head and Optimization

Figure 2 is the block-level architecture, and it consists of parallel paths, a quality-aware cross-gate, causal pooling and a calibrated forecast head. The work will be a white-background academic flow chart.

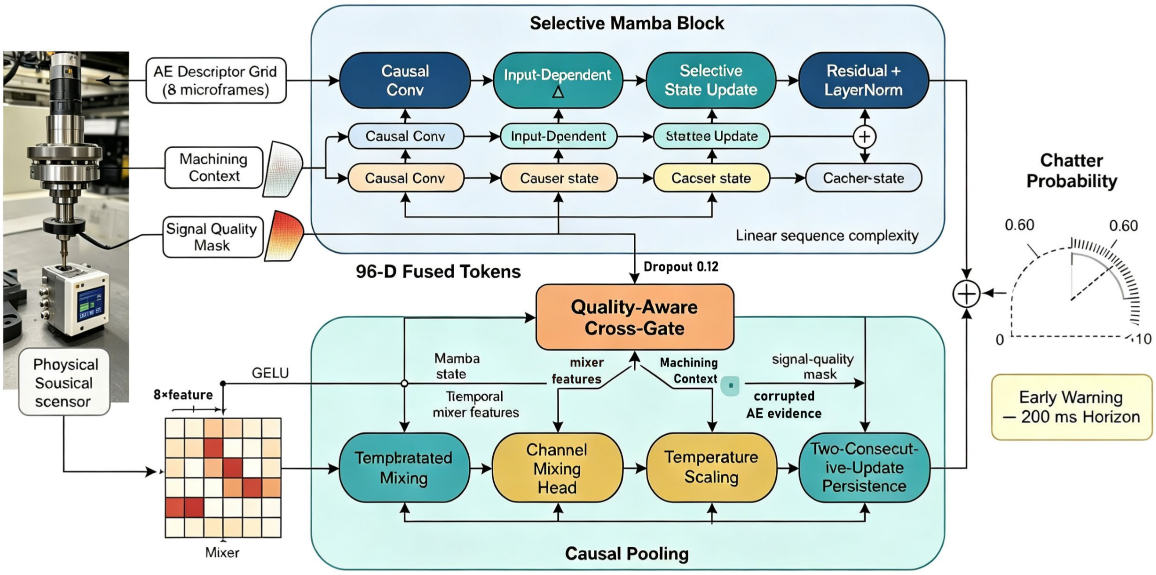

Figure 2. Dual-path Mamba-TSMixer architecture with selective state-space memory, temporal-channel mixing, and quality-aware cross-gated fusion.

A weighted focal objective addresses the short positive interval before each onset. A secondary lead-consistency term discourages probability oscillation when an event approaches.

$$\mathcal{L}_f = -\alpha(1-p_t)^{\gamma} y_t \log p_t - (1-\alpha)p_t^{\gamma}(1-y_t)\log(1-p_t) \quad \text{Eq.(13)}$$

The total objective includes calibration and smoothness penalties, with weights selected on grouped validation runs.

$$\mathcal{L} = \mathcal{L}_f + \lambda_s \sum_t |p_t - p_{t-1}| I_t + \lambda_c \mathrm{Brier}(\mathbf{p}, \mathbf{y}) \quad \text{Eq.(14)}$$

Inference uses the most recent cached state and emits a probability every 64 ms. Temperature scaling is fitted after training, and a chatter warning is issued only when two consecutive probabilities exceed 0.60. The fixed rule will be used for testing.

The four state-space blocks and the three mixer blocks are arranged as follows. Hidden Width is 96, State Dimension is 24, Temporal Expansion is two, and Channel Expansion is three. Dropout of 0.12 is used after each residual transformation. AdamW optimization starts at 3×10^{-4}, uses cosine decay, has five warm-up epochs, a batch size of 64, and gradient clipping of 1. Training stops when grouped-validation event F1 does not improve for 12 epochs. A class weight is calculated for each training fold and a focal exponent of 2 is set. These selections are fixed before the held-out periods are decoded.

Causality takes the following three forms. Signal windows end at the update time, temporal convolutions use left padding, and validation calibration never observes test results. The state cache is cleared before each cutting run and upon controller-reported tool changes. This reset policy does not carry evidence across physically distant operations. A compact diagnostic output stores the cross-gate mean, quality confidence and top feature-group contribution for each alarm to help the operator interface differentiate a high-confidence instability trajectory from a warning issued under degraded sensing.

Experimental Evaluation and Diagnostic Analysis

Testbed, Dataset, and Evaluation Protocol

In the experiment on the three-axis machining center, a 10mm four-flute carbide end mill and an aluminum workpiece were used. The Spindle-mounted AE sensor is relatively close to the front bearing housing. The 1,260 revolutions per minute of the spindle had an axial depth of 0.5-3.0 mm and a radial immersion ratio of 0.15-0.70. Four wear states were obtained by controlled cumulative cutting. Recent studies on machining-monitoring have

https://doi.org/10.64972/jaat.2024v2.341p33e:456-469

focused on run-level separation when temporal windows are highly correlated [25]; therefore, 70% of the complete runs were used to form the training set, 10% were grouped into validation sets, and 20% were used for testing. Two speed-depth combinations and the most worn tool state were excluded from the training set of unseen-regime tests.

Event F1 is a better index because window accuracy is increased by extended stable periods. Other supporting evidence includes the Area Under the Receiver Operating Characteristic Curve (AUROC), the area under the Precision-Recall Curve, false alarms per cutting minute, median warning lead, expected calibration error, and per-update latency. Event matching accepts one warning in the 0-400ms window before a labelled onset. This event-driven approach meets the requirement for alarm evaluation and does not use standalone windows [26]. Baselines are Random Forest, one-dimensional convolution, temporal convolutional network, gated recurrent unit, long short-term memory, Transformer, PatchTST and ungated TSMixer. Hyperparameters are chosen from the same group of validation budgets for fair time-series comparison with [27].

The test part has 252 runs, 96 chatter events and 3.7 h of active cutting. Stable time accounts for 78.4% of the decoded updates, the pre-onset positive horizon accounts for 8.9%, and post-onset or transition intervals excluded from early-warning scoring make up the rest. Onset annotation is carried out independently by two annotators using synchronous vibration spectra and audio playback, and disagreements exceeding one update are resolved based on the order map. Median absolute inter-annotator difference before adjudication is 32 ms. Confidence intervals are calculated by resampling whole runs 2,000 times, preserving the clustering of windows within an operation.

All baselines receive the same standardized descriptor grid and contextual variables; only the random forest is a flattened summary statistic. Their parameter budgets are limited to 2-7 million in some places. Sequence length, optimizer family, early stopping rule, and maximum training epochs are all shared. Transformer has causal masking and four attention heads; PatchTST uses non-overlapping temporal patches; recurrent networks have two stacked layers. Runtime is the descriptor assembly, model execution and probability calibration on an industrial PC with an eight-core processor. GPU timing is only shown as the auxiliary diagnostic information because the target controller path is CPU-based.

Predictive Accuracy, Early Warning, and Robustness

Figure 3(a) compares event F1 across nine models: random forest 0.781, one-dimensional convolution 0.842, temporal convolutional network 0.867, gated recurrent unit 0.871, long short-term memory 0.879, Transformer 0.891, PatchTST 0.905, TSMixer 0.907, and the proposed model 0.943. Figure 3(b) shows the corresponding AUROC values from 0.842 to 0.961 together with 95% bootstrap intervals. Figure 3(c) plots false alarms against median warning lead; Mamba-TSMixer reaches 184 ms at 0.18 false alarms per minute, whereas the strongest baseline reaches 151 ms at 0.31. The complete accuracy comparison appears in Figure 3. Deep temporal baselines are competitive [28], but the fused paths improve both event discrimination and operational alarm quality rather than trading one for the other.

At the event level, the proposed model finds 91 of the 96 chatter episodes, produces five missed events, and generates 14 unmatched warnings. TSMixer found 87 events and had 24 unmatched warnings; PatchTST found 86 events and had 22. Gradual onsets show the most improvement; only an increase of 1.6 dB in the AE envelope occurs during the first half of the forecast interval, but sideband regularity rises. Selective Memory accumulates this weak evidence over multiple updates. For abrupt onsets, all deep baselines perform more similarly because a single high-energy block already separates the classes. Precision-recall area is 0.914 for TSMixer and 0.951 respectively; therefore, the F1 gain is not limited to the selected threshold.

Bootstrap tests show that Bootstrap is 3.8 per cent higher than TSMixer in the F1 score, and the 95% confidence interval is 2.1 - 5.6. The interval excludes zero in 1,974 of the 2,000 run-level resamples. AUROC improvement is small because many stable windows are easy for all sequence models; the operational difference is concentrated near entry transitions and low-amplitude pre-onset periods. False Alarms are verified by cutting phase. Entry and exit account for 42%, chip re-cutting for 21%, rapid speed changes for 18%, and unexplained high-frequency bursts for 19%. Cross-gating reduces the entry-related alarm most effectively because the engagement and speed derivatives are contextually recognizable bursts.

https://doi.org/10.64972/jaat.2024v2.341p33e:456-469

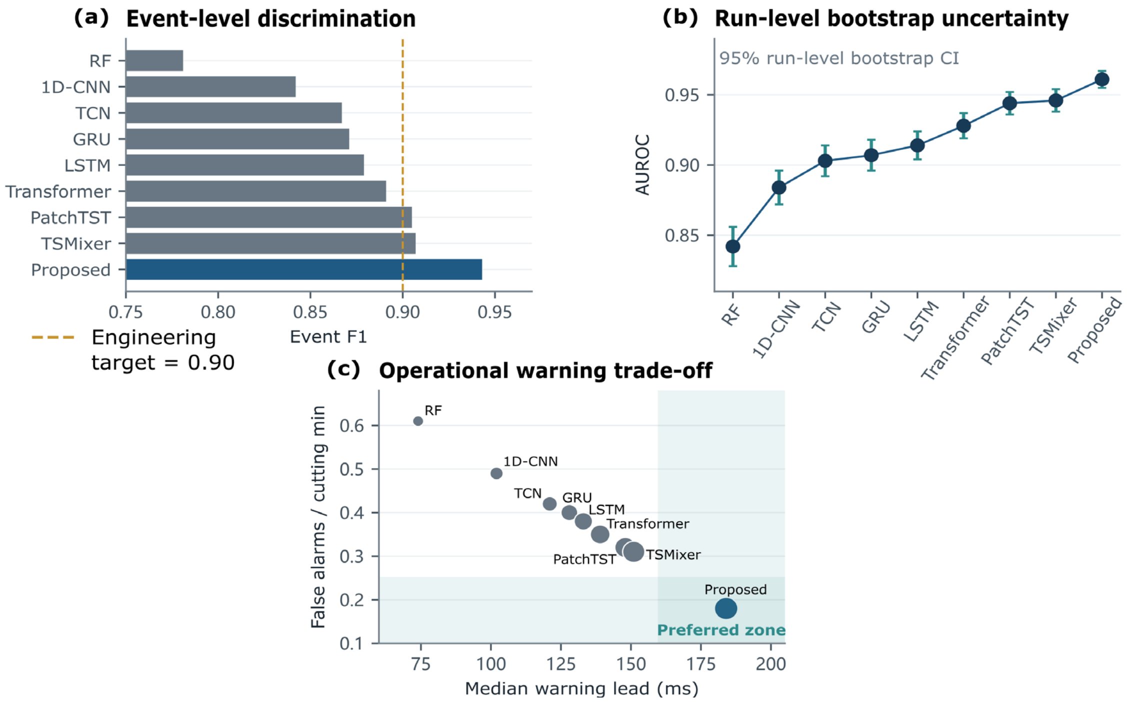

Figure 3. Overall prediction performance: (a) event F1 by model; (b) AUROC with 95% bootstrap confidence intervals; (c) false-alarm rate versus median warning lead.

Figure 4(a) tracks probability during a representative speed ramp: the proposed curve crosses the 0.60 threshold 192 ms before onset and remains above it, while Transformer oscillates twice and crosses only 96 ms early. Figure 4(b) presents an error-density ridge over five speed regimes; median timing error remains between −18 and 23 ms for Mamba-TSMixer but broadens beyond 70 ms for the recurrent baseline at 12,000 r/min. Figure 4 therefore connects sample-level probability evolution with regime-level timing stability. Speed-conditioned validation is important because distribution shift can otherwise be hidden by random window splits [29].

The range of the interquartile range for all detected events' warning lead is 151–231ms. Eleven alarms occur earlier than 300 ms; these are counted as detections, but their lead is clipped at the evaluation boundary to prevent a long stable fluctuation from affecting the median. Five late detections occur after the 200ms target but before annotated onset. They are still operationally useful but are treated as horizon errors in the timing analysis. The probability slope of the last three updates is 0.119 per update for true warnings and 0.037 for unmatched warnings; thus, the trajectory shape may support a future graded alarm policy. Speed-stratified event F1 is 0.951, 0.947, 0.944, 0.936, and 0.921 for the lowest to the highest regime. A drop to 12,000 r/min occurs alongside a general AE background energy reduction and shorter tooth-pass intervals; however, specificity remains over 0.95. Warning lead is reduced by 27ms in the same range. Using normalized sequential features of measured speed compared to using fixed frequency bands, the F1 value returns to 0.014, so synchronization is independent of the sequence architecture.

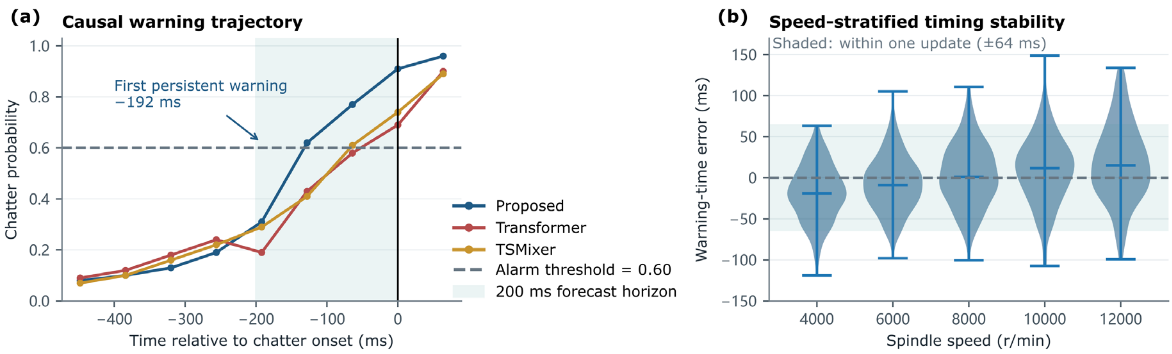

Figure 4. Temporal warning behavior: (a) probability trajectories, threshold, and true onset during a speed ramp; (b) warning-time error distributions across five spindle-speed regimes.

https://doi.org/10.64972/jaat.2024v2.341p33e:456-469

Robustness is summarized in Figure 5. Figure 5(a) shows event F1 under signal perturbations: clean 0.943, masked AE band group 0.912, 10% packet loss 0.901, −6 dB impulsive noise 0.884, and missing context channel 0.895. Figure 5(b) uses a polar profile to compare sensitivity, specificity, calibration, lead time, and noise tolerance; the proposed model encloses the largest balanced area. Figure 5(c) is the only heat map and reports F1 for five speeds by four wear states, ranging from 0.887 in the most difficult high-speed worn condition to 0.956 under moderate wear. Figure 5(d) presents calibration reliability: temperature scaling reduces expected calibration error from 0.064 to 0.021. Sensor-fusion reliability should be evaluated under partial corruption rather than only on clean recordings [30], and calibration matters when a probability controls a machine response [31].

Masking all AE features reduces F1 to 0.824; masking context variables reduces it to 0.895. Therefore, AE is still the main predictor, and context serves to disambiguate it. When an AE band group is missing, the gating will adjust its weight to favor envelope dynamics and orderly sidebands; its average AE contribution decreases from 0.68 to 0.54, but it will not drop to zero. At a packet loss of 20%, F1 is 0.876 and the median lead is 158ms. Packet-loss bursts longer than three updates are responsible for most of the misses, so controller buffering is still required even with feature masking augmentation.

Calibration bins have had about the same number of updates. Before scaling, the predictions in the range [0.7, 0.9] are overly confident by 6-9 percentage points. The temperature is 1.34; thus, the ranking metrics are unchanged. Brier score is 0.071 and 0.058, respectively. Due to unobserved wear, the expected calibration error has reached 0.034; although it is still lower than the unscaled in-distribution value, a threshold transfer warning should be issued after significant tool degradation.

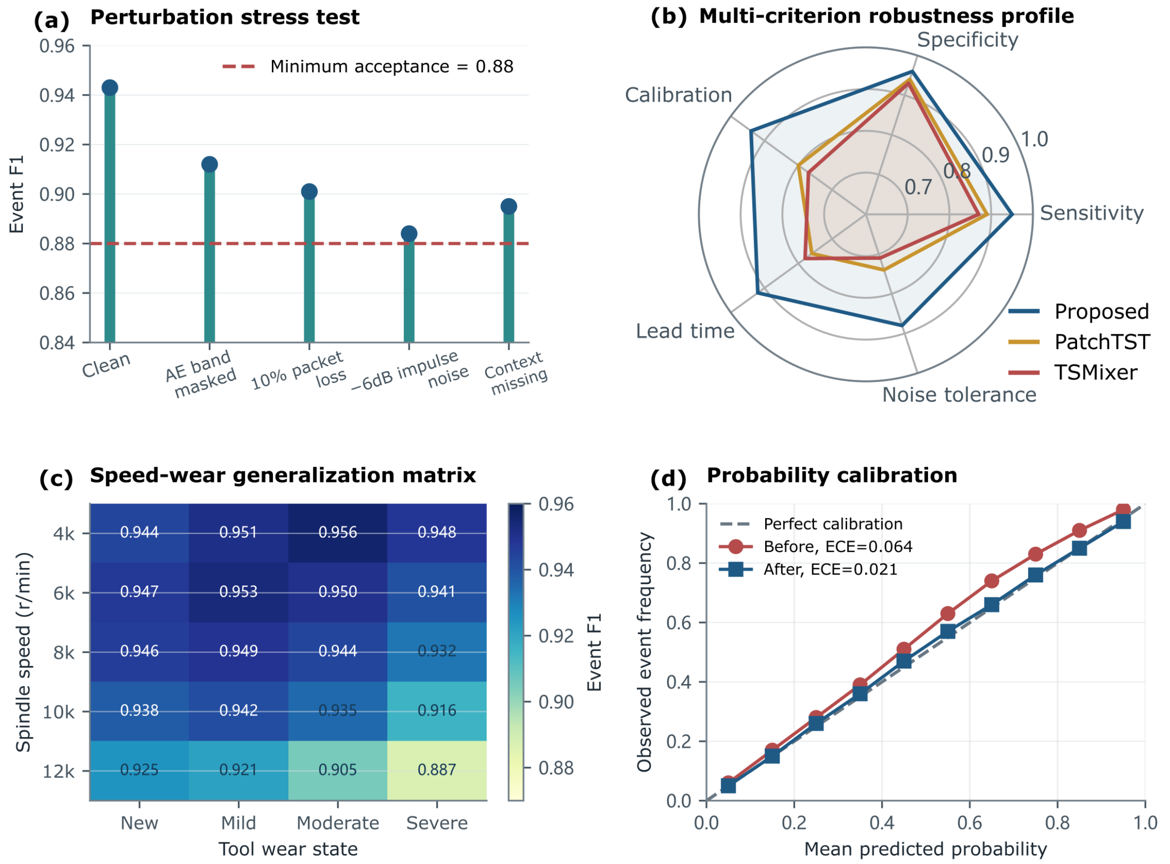

Figure 5. Robustness and confidence diagnostics: (a) perturbation stress test; (b) multi-criterion radar profile; (c) speed-by-wear F1 heat map; (d) reliability diagram before and after calibration.

Ablation, Computational Cost, and Engineering Interpretation

Figure 6(a) reports the ablation waterfall. Removing cross-gating reduces F1 by 0.021, replacing selective state-space memory with fixed recurrence reduces it by 0.028, removing the mixer reduces it by 0.017, and using AE

https://doi.org/10.64972/jaat.2024v2.341p33e:456-469

energy alone reduces it by 0.049. Figure 6(b) plots parameter count against latency: the proposed model uses 2.84 million parameters and 1.8 ms per update, compared with PatchTST at 4.91 million and 3.4 ms and Transformer at 6.20 million and 4.7 ms. Figure 6(c) is a violin distribution of latency over 2,000 updates; the 99th percentile is 2.35 ms and remains far below the 64 ms update interval. Efficient sequence models are valuable only when their runtime advantage survives the complete preprocessing and forecast path [32]. The Figure 6 evidence indicates that the gain is not explained by a larger model or relaxed timing budget.

Pairwise ablation shows that the two paths interact. Both the mixer and cross-gate were removed, and F1 dropped to 0.886; it was a greater loss than either isolated removal. Reducing the state dimension from 24 to 8 saves 0.22 ms but loses 0.013 F1; increasing it to 48 adds 0.39ms for only an improvement of 0.003 F1. Context length 256 ms has an F1 score of 0.921, context length 512 ms is 0.943, and context length 768 ms is 0.944. Therefore, the selected configuration is relatively close to the knee of the accuracy-cost curve.

Descriptor computation takes 0.74 ms of the 1.8 ms median update, the neural network is 0.91 ms, and calibration plus alarm persistence is 0.15 ms. On a low-power four-core edge processor, the median latency is 4.9 ms and the 99th-percentile is 6.8 ms. Both are still less than one-tenth of the update interval. Peak working memory is 46 MB and includes the streaming AE buffer. Caching the state reduces computation by 37 per cent compared to re-evaluating the whole 512 ms sequence.

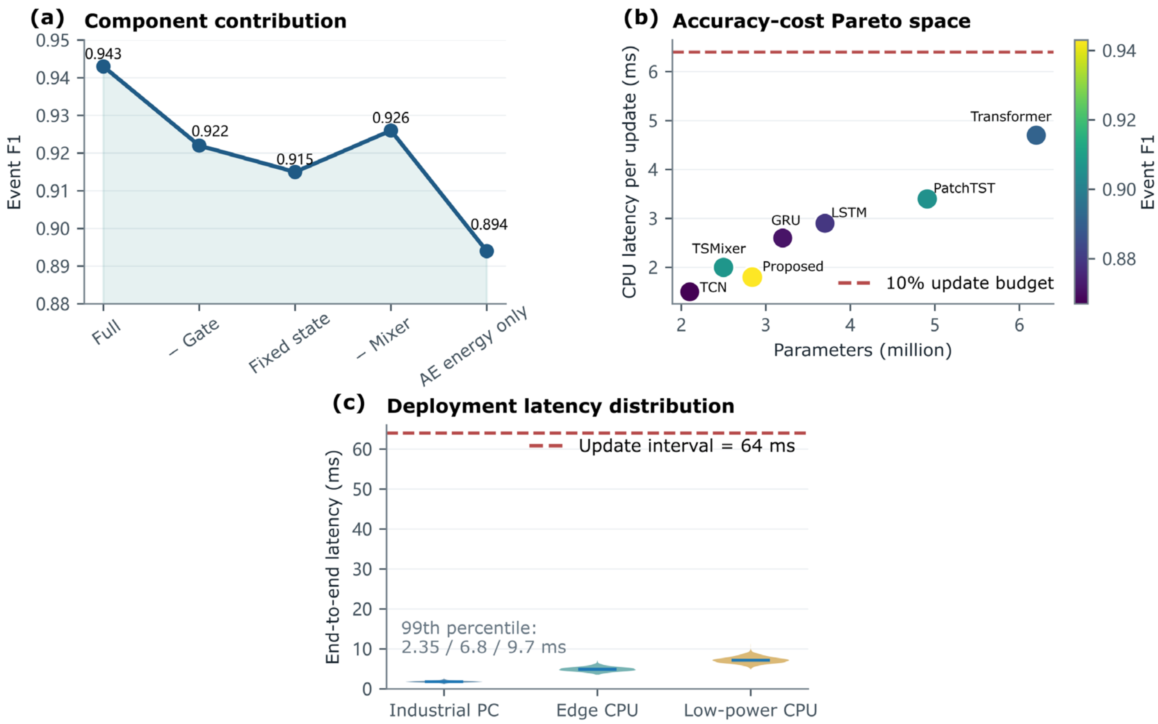

Figure 6. Component and runtime analysis: (a) ablation waterfall in event F1; (b) parameter-latency Pareto scatter; (c) inference-latency distributions across deployment targets.

Figure 7(a) visualizes the two-dimensional embedding of test windows, with stable, transition, and chatter clusters; transition samples form a narrow bridge rather than being absorbed into the stable class. Figure 7(b) gives feature-group permutation importance: AE sidebands contribute 0.031 F1, envelope dynamics 0.026, impulsiveness 0.019, spindle context 0.017, and signal quality 0.011. Figure 7(c) uses a normalized stacked horizontal bar chart to stratify 214 difficult events by disturbance source. Entry/exit transients account for 82 events, of which 58 (71%) are resolved by gated fusion; severe wear accounts for 51 events, with 32 (63%) resolved; speed ramps account for 44 events, with 30 (68%) resolved; and contamination accounts for 37 events, with 21 (57%) resolved. Absolute counts, within-source percentages, and category totals are displayed together to support direct comparison. Interpretation methods should be treated as diagnostic probes rather than causal proof [33]. Nonetheless, the three views in Figure 7 consistently show that the network combines high-frequency AE evidence with operating context instead of relying on absolute amplitude.

Embedding neighborhoods are quantified rather than interpreted only by appearance. Silhouette score for the three states is 0.46 for Mamba-TSMixer, 0.34 for TSMixer, and 0.29 for PatchTST. Transition-to-chatter distance decreases monotonically over the final four pre-onset updates in 79 of 96 events. Permuting sidebands harms high-speed runs most, while permuting envelope dynamics harms low-speed gradual onsets. The complementary pattern agrees with the designed division between frequency-conditioned and local temporal evidence.

Manual inspection of the five missed events finds three under severe wear with continuous broadband emission, one during simultaneous speed and immersion change, and one with a tachometer dropout. Of the 14 false alarms, six precede heavy entry impacts, four occur during chip re-cutting, two coincide with coolant bursts, and two have no clear external cause. The quality gate is low for the tachometer-dropout miss but not for broadband wear, indicating that future adaptation should model slow background drift separately from acute sensor corruption.

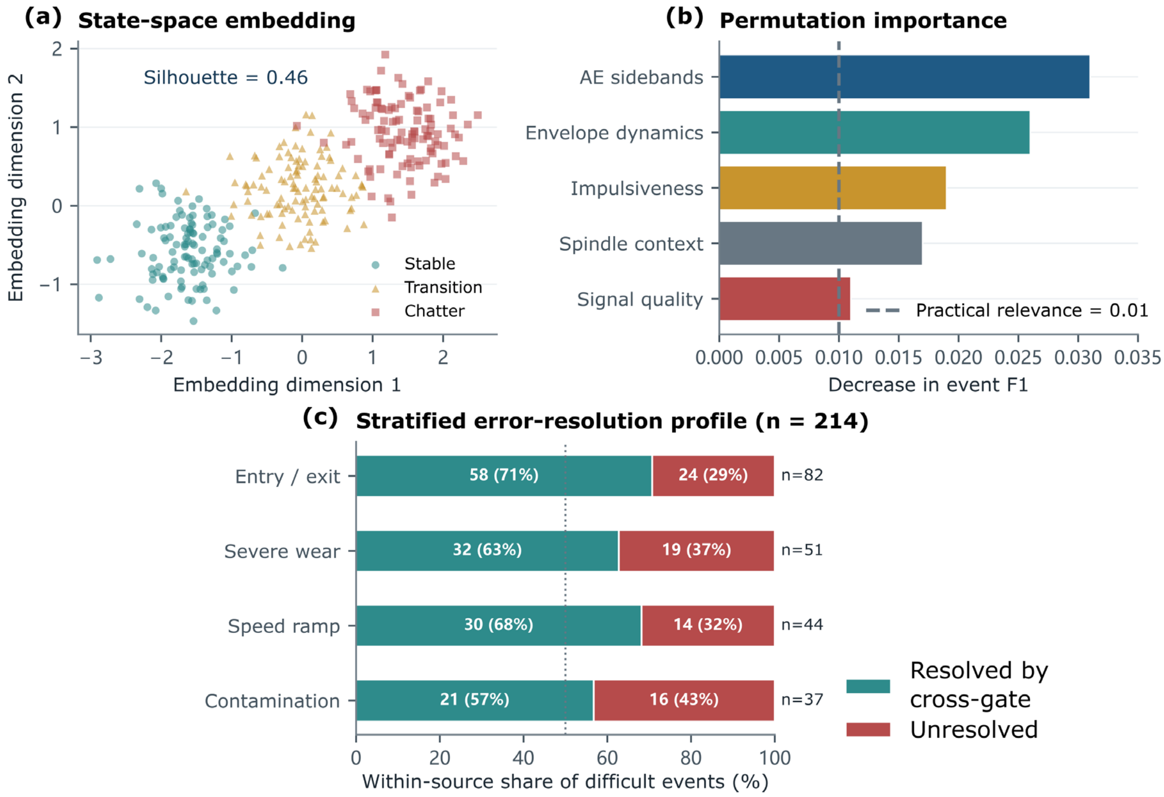

Figure 7. Representation and error interpretation: (a) test-window embedding by machining state; (b) permutation importance of five feature groups; (c) normalized stacked error-resolution profile across four disturbance sources, with absolute counts and within-source percentages.

The operational threshold was also tested over 0.45–0.75. Event F1 remained above 0.925 between 0.55 and 0.66, while warning lead declined from 211 to 146 ms as the threshold increased. A value of 0.60 therefore provides a stable operating region rather than a single tuned optimum. The unseen-regime subset produced 0.918 F1 and 0.936 AUROC, compared with 0.949 and 0.966 on familiar regimes. Domain shift remains measurable, but its magnitude is smaller than for PatchTST, whose F1 falls from 0.916 to 0.872. Evaluation across operating domains is essential for machine-monitoring deployment [34]. Repeating training with five seeds yields a 0.943 mean F1 and 0.006 standard deviation. Bootstrap intervals, run-grouped splitting, and fixed alarm persistence reduce the chance that near-duplicate windows or favorable threshold selection explain the result [35].

The other is a change in the prediction period while keeping the same cause-and-effect conditions. At 100 ms, event F1 is 0.956 and median warning lead is 92 ms; at 200 ms, they are 0.943 and 184 ms; at 300 ms, they are 0.918 and 261 ms; and at 400 ms, they are 0.887 and 337 ms. A long-horizon exposure has weaker and less regular precursors; therefore, a decline is expected. A 200ms delay is considered long enough for the controller to recognize the alarm and request a parameter modification without causing frequent network classifications of stable signals. Re-training is required for each horizon; otherwise, shifting the threshold earlier in the 200ms model will reduce precision because the probability has not been adjusted for that decision boundary.

Repeatability is shown for three sensor remounts at the same nominal preload. If remapping normalization is not performed, the range of F1 is from 0.901 to 0.944, because the shell transmission path will change the envelope amplitude. Robust training-set scaling and band ratio narrowing reduce the range to 0.929-0.946. Only fine-tune the input projection for the ten stable and five chatter runs from a remounted sensor, and the lowest result is 0.938. Limited calibration does not alter the state-space or mixer blocks, and it completes in less than three minutes at the training workstation. It offers a practical commissioning path and keeps the general representation.

Surface measurements supply an external engineering check on alarm meaning. Runs warned with probability above 0.8 have median arithmetic roughness of 3.6 $\mu\mathrm{m}$ if allowed to continue, compared with 1.1 $\mu\mathrm{m}$ for confidently stable runs under matched feed per tooth. Borderline warnings between 0.6 and 0.7 have a wider 1.4–2.8 $\mu\mathrm{m}$ range, consistent with their mixed transition state. These measurements are not used as labels, because roughness integrates effects beyond chatter, but they show that the forecast score aligns with a consequence relevant to production quality.

The alarm logic does not need to alter the machine control kernel. The monitoring process reads timestamped speed and engagement variables, updates the AE ring buffer, computes descriptors, advances the cached model state, and publishes probability plus quality confidence. If the two updates exceed the threshold and the quality confidence is above 0.5, then an actionable warning is displayed in the interface; otherwise, a degraded-sensing advisory is recorded. Therefore, a missing tachometer will not be mistakenly regarded as a normal cut. A controller can be added to map warnings to feed hold, speed adjustment, or operator notification, but the current experiment only tests prediction and does not claim to achieve closed-loop stability.

Error Costs Differ According to Manufacturing Processes. Finish: Omission of an event will damage the finished surface; a false alarm will only reduce throughput. Roughing can be relatively conservative at times, but repeated entry alarms are inconvenient. Using validation probabilities, a cost ratio of five for misses to false alarms is chosen to select a threshold of 0.57, and 0.947 recall is achieved with 0.203 false alarms per minute. Equal costs choose 0.64, and the results are 0.918 recall and 0.159 false alarms per minute. Report the scope to make the model suitable for production without re-optimizing the weights.

Conclusion

Develop a causal Mamba-TSMixer model in this paper to transform spindle acoustic-emission signals and machining context into an early chatter warning. Selective state-space memory is gradual pre-onset evolution; temporal-channel mixing resolves local descriptor interactions; and quality-aware cross-gating limits the impact of impulsive or missing measurements. The protocol for the accompanying signal, horizon definition, grouped evaluation and deployment-oriented metrics form a complete path from high-frequency sensing to a controller-ready probability.

At present, this work is limited to a single machine platform, one workpiece family, and a sensor mounting arrangement that did not change during data collection. The start label is based on a consensus reference rather than a directly observable physical moment, and the model does not adapt explicitly after significant spindle maintenance or structural modification. The above constraints rule out a claim of general transfer, or else the hold-out group test shows significant performance fluctuation.

Future work will explore continuous calibration for all machines, uncertainty-aware adaptation after tool replacement, and fusion with motor-current or spindle-displacement channels where available. Another way is to connect forecast confidence with constrained feed or speed adjustment, and thus the monitoring model can support graded intervention instead of a binary alarm. Public multi-machine benchmarks that coordinate high-frequency emissions and controller context would also increase the strictness of cross-platform evaluation. Such studies should also report the duration of alarms, time uncertainty, computational jitter, and sensor quality status, so that the predicted improvements can be realized as safe and auditable machine responses. Expand the scope of validation to cover all shifts.

Author Contributions

Karolina Tatiana Kuśowa contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Ludwik Lupa contributes to data collection, draft preparation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

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

Institutional Review Board Statement

Not applicable.

References

[1] Li, M., Huang, D., & Yang, X. (2021). Chatter stability prediction and detection during high-speed robotic milling process based on acoustic emission technique. The International Journal of Advanced Manufacturing Technology, 117(5-6), 1589-1599. https://doi.org/10.1007/s00170-021-07844-3

[2] Sestito, G. S., Venter, G. S., Ribeiro, K. S. B., Rodrigues, A. R., & da Silva, M. M. (2022). In-process chatter detection in micro-milling using acoustic emission via machine learning classifiers. The International Journal of Advanced Manufacturing Technology, 120(11-12), 7293-7303. https://doi.org/10.1007/s00170-022-09209-w

[3] Deng, C., Zhou, W., Yang, K., Huang, Z., & Tang, Q. (2020). Robust Chatter Stability Prediction of the Milling Process considering Uncertain Machining Positions. Mathematical Problems in Engineering, 2020, 1-15. https://doi.org/10.1155/2020/1398957

[4] Bhakar, G., Khandagale, P., Sonawane, H., & Joshi, S. S. (2021). Strategy development for chatter-free milling of Ti-6Al-4V thin-walled surfaces using stability region diagram (SRD). Machining Science and Technology, 25(6), 899-929. https://doi.org/10.1080/10910344.2021.1903925

[5] Itoh, M., Hayasaka, T., & Shamoto, E. (2020). High-efficiency smooth-surface high-chatter-stability machining of thin plates with novel face-milling cutter geometry. Precision Engineering, 64, 165-176. https://doi.org/10.1016/j.precisioneng.2020.03.024

[6] Wahid, M. K., Kasno, M. A., Ahmad, M. N., Osman, M. H., Maidin, N. A., Rahman, M. H. A., & Firdaus, H. M. S. (2020, April). Improvement of chatter stability in milling machining using neoprene rubber sheet as damping element. In Journal of Physics: Conference Series (Vol. 1529, No. 4, p. 042033). IOP Publishing. https://doi.org/10.1088/1742-6596/1529/4/042033

[7] Chen, K., Zhang, X., Zhao, Z., Yin, J., & Zhao, W. (2021). Milling chatter monitoring under variable cutting conditions based on time series features. The International Journal of Advanced Manufacturing Technology, 113(9-10), 2595-2613. https://doi.org/10.1007/s00170-021-06746-8

[8] Sio-Sever, A., Leal-Muñoz, E., Lopez-Navarro, J., Alzugaray-Franz, R., Vizan-Idoipe, A., & de Arcas-Castro, G. (2020). Non-Invasive Estimation of Machining Parameters during End-Milling Operations Based on Acoustic Emission. Sensors, 20(18), 5326. https://doi.org/10.3390/s20185326

[9] NIIZEKI, H. (2021). Prediction of chatter vibration in milling. The Proceedings of The Computational Mechanics Conference, 2021.34(0), 066. https://doi.org/10.1299/jsmecmd.2021.34.066

[10] Tehranizadeh, F., & Budak, E. (2022). Special Milling Tools to Suppress Self-Excited Chatter Vibrations in Milling. IFAC-PapersOnLine, 55(36), 115-120. https://doi.org/10.1016/j.ifacol.2022.11.343

[11] Huang, C., Yang, W. A., Cai, X., Liu, W., & You, Y. (2020). An Efficient Third-Order Full-Discretization Method for Prediction of Regenerative Chatter Stability in Milling. Shock and Vibration, 2020, 1-16. https://doi.org/10.1155/2020/9071451

[12] Ji, Y., Wang, L., Song, Y., Wang, H., & Liu, Z. (2022). Investigation of robotic milling chatter stability prediction under different cutter orientations by an updated full-discretization method. Journal of Sound and Vibration, 536, 117150. https://doi.org/10.1016/j.jsv.2022.117150

[13] Shaik, J. H., Mutra, R. R., Galal, A. M., & Tanneeru, S. R. (2022). Investigation of chatter prediction in end milling using morphological studies on aluminum alloy (Al6061). Noise & Vibration Worldwide, 53(4-5), 233-243. https://doi.org/10.1177/09574565221093232

[14] Zhao, N., Su, Y., Wang, S., Xia, M., & Liu, C. (2022). Chatter Detection in Variable Cutting Depth Side Milling Using VMD and Vibration Characteristics Analysis. Electronics, 11(22), 3779. https://doi.org/10.3390/electronics11223779

[15] Gao, J., & Altintas, Y. (2020). Chatter stability of synchronized elliptical vibration assisted milling. CIRP Journal of Manufacturing Science and Technology, 28, 76-86. https://doi.org/10.1016/j.cirpj.2019.11.006

[16] Vashisht, R. K., & Peng, Q. (2021). Fractional Calculus-Based Energy Efficient Active Chatter Control of Milling Process Using Small Size Electromagnetic Actuators. Journal of Vibration and Acoustics, 143(1), 011005. https://doi.org/10.1115/1.4047703

[17] Emami, M., & Karimipour, A. (2021). Theoretical and experimental study of the chatter vibration in wet and MQL machining conditions in turning process. Precision Engineering, 72, 41-58. https://doi.org/10.1016/j.precisioneng.2021.04.006

[18] Beri, B., Meszaros, G., & Stepan, G. (2021). Machining of slender workpieces subjected to time-periodic axial force: stability and chatter suppression. Journal of Sound and Vibration, 504, 116114. https://doi.org/10.1016/j.jsv.2021.116114

[19] Jeong, K., Seong, Y., Jeon, J., Moon, S., & Park, J. (2022). Chatter Monitoring of Machining Center Using Head Stock Structural Vibration Analyzed with a 1D Convolutional Neural Network. Sensors, 22(14), 5432. https://doi.org/10.3390/s22145432

[20] Hesabi Hesari, A., Moradi, H., & Movahhedy, M. R. (2020). Modeling and active control of chatter vibrations of piezoelectric stacked machining arm in micro-turning process. Meccanica, 55(9), 1707-1731. https://doi.org/10.1007/s11012-020-01210-5

[21] Yang, K., Wang, G., & Cong, J. (2022). Milling chatter monitoring based on sparse representation and image similarity measurement. Insight - Non-Destructive Testing and Condition Monitoring, 64(3), 146-154. https://doi.org/10.1784/insi.2022.64.3.146

[22] Zhao, X., Zheng, L., Liu, X., & Cao, Y. (2020). Chatter stability prediction for multi-robots collaborative milling system. Procedia CIRP, 93, 856-861. https://doi.org/10.1016/j.procir.2020.03.138

[23] Mishra, R., & Singh, B. (2022). SB-LMD based online monitoring of tool chatter detection in milling process. Materials Today: Proceedings, 56, 3276-3284. https://doi.org/10.1016/j.matpr.2021.09.489

[24] Lin, L., He, M., Wang, Q., & Deng, C. (2021). Chatter Stability Prediction and Process Parameters' Optimization of Milling Considering Uncertain Tool Information. Symmetry, 13(6), 1071. https://doi.org/10.3390/sym13061071

[25] Zhu, L., & Liu, C. (2020). Recent progress of chatter prediction, detection and suppression in milling. Mechanical Systems and Signal Processing, 143, 106840. https://doi.org/10.1016/j.ymssp.2020.106840

[26] Mishra, R., & Singh, B. (2022). Prediction of milling chatter using SBLMD-ANN. Journal of Mechanical Science and Technology, 36(2), 877-882. https://doi.org/10.1007/s12206-022-0135-5

[27] Dun, Y., Zhu, L., & Wang, S. (2020). Multi-modal method for chatter stability prediction and control in milling of thin-walled workpiece. Applied Mathematical Modelling, 80, 602-624. https://doi.org/10.1016/j.apm.2019.12.003

[28] Yamato, S., Nakanishi, K., Suzuki, N., & Kakinuma, Y. (2021). Development of Automatic Chatter Suppression System in Parallel Milling by Real-Time Spindle Speed Control with Observer-Based Chatter Monitoring. International Journal of Precision Engineering and Manufacturing, 22(2), 227-240. https://doi.org/10.1007/s12541-021-00469-2

[29] Li, W., Wang, L., & Yu, G. (2022). Chatter prediction in flank milling of thin-walled parts considering force-induced deformation. Mechanical Systems and Signal Processing, 165, 108314. https://doi.org/10.1016/j.ymssp.2021.108314

[30] Chen, D., Zhang, X., Zhao, H., & Ding, H. (2021). Development of a novel online chatter monitoring system for flexible milling process. Mechanical Systems and Signal Processing, 159, 107799. https://doi.org/10.1016/j.ymssp.2021.107799

[31] Liu, B., Liu, C., Yu, X., Zhou, Y., & Wang, D. (2022). Prediction, detection, and suppression of regenerative chatter in milling. Advances in Mechanical Engineering, 14(10), 16878132221129746. https://doi.org/10.1177/16878132221129746

[32] Totis, G., & Sortino, M. (2020). Polynomial Chaos-Kriging approaches for an efficient probabilistic chatter prediction in milling. International Journal of Machine Tools and Manufacture, 157, 103610. https://doi.org/10.1016/j.ijmachtools.2020.103610

[33] Wang, X., Song, Q., & Liu, Z. (2020). Precise chatter monitoring of thin-walled component milling process based on parametric time-frequency transform method. Journal of Materials Processing Technology, 283, 116712. https://doi.org/10.1016/j.jmatprotec.2020.116712
[34] Li, W., Wang, L., Yu, G., & Wang, D. (2021). Time-varying dynamics updating method for chatter prediction in thin-walled part milling process. Mechanical Systems and Signal Processing, 159, 107840. https://doi.org/10.1016/j.ymssp.2021.107840
[35] Wei, X., Miao, E., & Ye, H. (2022). Analytical Prediction of Three Dimensional Chatter Stability Considering Multiple Parameters in Milling. International Journal of Precision Engineering and Manufacturing, 23(7), 711-720. https://doi.org/10.1007/s12541-022-00645-y