UNIVERSITY OF WEST ATTICA

eISSN:3136-7658

Applied Manufacturing Intelligence Journal

Vol. 5, No. 4, October, 2026



S4D Mixer Multisource Fusion for Carbon Fiber Tow Defect Recognition

Jindřich Kopecký[1,*]

[1] *Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, 16627, Czechia*

Received 05 June 2026; Revised 11 October 2026; Accepted 12 October 2026; Published 16 October 2026

Abstract

Accurate defect detection in carbon fiber tow requires an inspection system that can observe surface appearance, geometric displacement, and mechanical disturbance along the same moving material coordinate. This study proposes an S4D-Mixer multisource fusion model that aligns optical line scans, laser profilometry, and tension signals on an encoder-indexed grid, processes each source with diagonal state-space layers, and combines the resulting tokens through a reliability-aware gate and residual channel mixer. The model is designed to preserve local defect boundaries while retaining causal context across extended tow disturbances. A constructed benchmark with six tow states is used to examine recognition behavior, calibration, missing-source robustness, and latency assumptions. In the illustrative comparison, the complete model reaches 96.7% macro F1 and 96.67% accuracy, exceeds fixed fusion under degraded sensing by 4.0-7.2 percentage points, and keeps selective risk below 1.3% at 80% coverage. These values are not presented as production measurements; they provide a transparent numerical protocol for evaluating the proposed inspection design before physical trials. The results indicate that source-quality gating can reduce the influence of saturated images, missing profile returns, and tension spikes, while S4D processing supports consistent sequence-level recognition. The framework offers a practical route for future in-line carbon tow inspection once validated with independently collected production data, hardware timing measurements, and lot-separated annotation.

Keywords: Carbon Fiber Tow, Defect Recognition, Diagonal State Space Models, Multisource Fusion, Sensor Reliability

1. Introduction

Research on the defects of carbon fiber tows has classified them based on the spatial characteristics of filament breakage, protruding fuzz, splitting and local distortion [1]. These local states affect downstream spreading, impregnation and composite uniformity [2]. The unresolved inspection gap is that these defects do not appear as isolated image objects only; they evolve along a moving tow and interact with illumination and bundle geometry [3]. They may also be accompanied by mechanical disturbances that are only weakly visible in a single optical frame [4].

Optical imaging is an attractive starting point because it provides direct evidence of exposed filaments and bundle edges, but the visibility of a defect remains conditional on resolution, contrast, and occlusion. Methods for textile and fiber inspection have used learned visual features to accommodate variations that are difficult to describe with fixed thresholds [5]. Such methods can retain fine texture, yet a short image crop may not reveal whether a narrow gap is an isolated fluctuation or part of a continuing split. Geometry based inspection introduces height and width information that can help distinguish a surface brightness changes from a physical displacement [6]. A height profile can nevertheless lose detail where reflectance is weak or the optical return is unstable. Tension measurements provide a different view

* Corresponding author: kopecky.j@fel.cvut.cz

https://doi.org/10.68406/amij.2026.vol5iss4np2:14-26

➢ 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.

of transport behavior and can identify a response that extends beyond the visible defect interval, although their spatial resolution is limited by the sampling rate and the measurement chain. Multisensor inspection research treats complementary sensing as a means of improving observability when one source has an ambiguous response [7]. Complementarity does not imply that every source is useful at every instant. Fusion studies also examine the adverse effect of combining misaligned or corrupted inputs as if their reliability were equal [8]. A recognition model for carbon fiber tows must therefore account for both the additional evidence and the additional failure modes introduced by multiple instruments. The engineering issue is not simply whether three sources outperform one on a clean benchmark; it is whether the representation continues to support sensible decisions when speed changes, a channel drops out, or a local disturbance affects only one instrument.

The temporal model determines how these heterogeneous observations are connected along the material path. Research on diagonal structured state-space models has proposed an efficient way to represent long-range dependencies without storing every pairwise token interaction, which is useful when a defect extends over many adjacent tow positions [9]. In the present setting, slowly decaying modes can preserve continuity across an extended disturbance, while direct and residual channel-mixing paths are needed to retain short boundary events that identify the beginning or end of a defect [10].

To connect the complementary observations along a moving tow, this study develops an S4D-Mixer architecture that integrates synchronous optical line scans, laser profile measurements and tension information. The paper first defines the acquisition representation and annotation contract, then presents the causal S4D-Mixer fusion model, evaluates the design through a constructed numerical benchmark, and finally discusses limitations and validation requirements for future production deployment.

2. Related Work

2.1. Appearance and Geometry Based Inspection

The first type of descriptors used for visual defect recognition are designed descriptors. Descriptor-based methods describe edges, textures or local intensity variations, and their decision rules are relatively simple to learn. Research on machine vision for fibrous materials has shown that such descriptors can be used to process regular surface patterns and defect shapes [11]. Their deficiency is that the descriptor response is affected by imaging conditions other than those of material quality. A brighter filament does not need to be due to a serious defect, and a weakly contrasted split may be a regular bundle spread. Learned convolutional representations can combine local structures of different sizes and reduce the number of rules needed for various appearances. Work on deep visual inspection investigates this advantage and also identifies the dependence of reported performance on annotation and acquisition conditions [12]. The above methods have good local evidence but cannot automatically connect events in multiple windows. Geometry provides a reference signal, and displacement, width, height, etc., describe changes in actual shape, not brightness. Profile-based inspection research has investigated the calibration and sampling requirements for interpreting the results of this measurement [13]. Combining appearance and geometry introduces a correspondence problem: measurements from separated instruments refer to the same material only after transport delay and position registration have been resolved. A model trained on incorrectly paired observations can learn associations that disappear when speed or instrument spacing changes. Existing work therefore motivates richer sensing while leaving the alignment contract and the consequences of unreliable profile returns as important questions for continuous tow inspection.

2.2. Long Sequence Models and Feature Mixing

Temporal Convolutional Networks use finite filters to process sequences and increase the effective context by stacking or dilating these filters. They are relatively simple and suitable for inspection streams; therefore, they should be used together. Temporal model research has explored the trade-off between a long receptive field and the preservation of short-term dependencies [14]. Attention-based models are also ways to connect the individual observations, and their computation characteristics vary according to the length of the sequence and how attention is implemented. Therefore, the comparison of inspections should distinguish between the effects of time and the effects of model size. Diagonal State Space Models are another type of accumulation and release mechanism for information at different scales in stable latent modes. Its suitability for tow inspection is a hypothesis about the length and durability of evidence, not the result of the architecture name. Another way to address this is to keep the features of surface texture, geometry and transport separate and process them in independent streams within the model. Literature on Multilayer Perceptron Mixing studies how to combine representations after token processing [15]. A combination of state-space dynamics and feature mixing has thus resulted in a feasible division of labor. Changes in materials are called long-term dynamics, and a combination of complementary source features is a mixing. The problems of application are causality, state-reset boundaries, and distinguishing between a real material change and a discontinuity due to altered acquisition conditions. The problems listed above need to be further specified according to the inspection and the data available at the time of prediction.

2.3. Fusion Reliability and Selective Recognition

Multi-source fusion can occur before feature extraction, among the modality encoders, or after the construction of separate decisions. Early fusion has a weak correlation and lacks coordinates and scales. Decision fusion is relatively easy to modularize, but some information that could help resolve the ambiguity of observation is lost. Research on uncertainty-aware sensor fusion studies intermediate representations that incorporate confidence or reliability of each source [16]. Reliability is not the same as class confidence; a sensor can produce a clean measurement that fails to distinguish between the two classes, or a corrupted measurement that is highly predictive because of a confounding pattern in the training set. If there is optical channel saturation, an invalid return profile or an intermittent offset of the tension sensor will be even worse. Lack of source training will reduce sensitivity to changes in availability, but the pattern of missingness must still be reported. The mean scores of some patterns are relatively low. Selective Recognition adds a decision step to route poorly supported predictions for manual review. The use cases depend on both the errors in the automatically retained cases and the proportion of production material that needs extra handling. Currently, the only deficiencies in the above ways are the lack of an incentive system for weighting and evaluation, no addition of a physical quality index, and no independent test split from the model. All of the above factors need to be considered in a study of tow recognition, otherwise the reliability of the fused prediction cannot be guaranteed.

3. Inspection Representation and Acquisition Workflow

3.1. Material Coordinate and Sensor Alignment

The proposed station observes one moving tow through an optical line scan camera, a laser profilometer, and a tension sensor. These instruments record exposed filaments, local height and width changes, and mechanical loading behavior from different physical viewpoints. Encoder position is used as the common material-coordinate reference so that measurements collected at different acquisition times can be mapped back to the same tow segment rather than merely compared by timestamp.

For source m, the raw measurement stream is mapped from acquisition time to encoder position and sampled on a common material grid. The representation uses a fixed spatial increment, Δx, and an instrument offset, d_m. Linear interpolation is suitable for smooth mechanical samples, whereas image lines and height profiles require resampling rules that preserve boundaries and invalidity information. The coordinate alignment equation expresses this mapping without assuming that the instruments observe the same material simultaneously:

$$u_k^{(m)} = \mathcal{R}_m\{s^{(m)}(t), x_t - d_m\}\big|_{x=k\Delta x} \tag{1}$$

Here, $s^{(m)}(t)$ denotes the raw source measurement, x_t is encoder position, and $\mathcal{R}_m$ is the source-specific resampling operator. The output $u_k^{(m)}$ represents the aligned measurement at material position $k\Delta x$. Resampling never crosses run boundaries or fills prolonged outages with apparently valid content. An availability mask and the age of the nearest genuine sample accompany each token. A 0.5 mm grid and 256 positions define a 128 mm context independent of transport speed. Instrument bandwidth and encoder accuracy must support this grid. Lot identifiers and original coordinate intervals enable an audit of overlap between nominally different windows.

3.2. Modality Encoders and Quality Descriptors

Modality Encoders Preserve Native Measurement Structure Before Fusion. An optical encoder performs a shallow spatial convolution of the adjacent camera lines and then pools over two widths to obtain width and edge descriptors. A profilometer encoder obtains the height, validity and cross-toe slope, and excludes invalid returns from normalization. A Tension Encoder acquires a phase-locked signal and provides local difference and interpolation angle data. A short causal convolution learns short-term changes. They are the same size, but the quality and explanatory power of the streams are not.

Quality descriptors measure source-specific reliability before fusion. For the optical stream, they include saturation, contrast, usable pixels and boundary coverage; for the laser profile, they include valid profile returns and missing samples; for the tension stream, they include clipping, abrupt change and measurement agreement with neighboring windows.

The workflow of acquisition and representation is shown in Figure 1, and it starts with the moving tow and ends at aligned source tokens, validity masks, and lot-linked labels. The four main branches are: Physical Sensing, Coordinate Registration, Quality Estimation and Annotation. Therefore, in the case of a recognition failure, one can determine whether it is due to missing input, a registration error, or a label problem. A division point for the calculation of training statistics is also shown in the flowchart; otherwise, normalization across all lots would leak information from the evaluation set. Although the architecture is for a variable number of sources, only the three instruments mentioned above are included in the current specification.

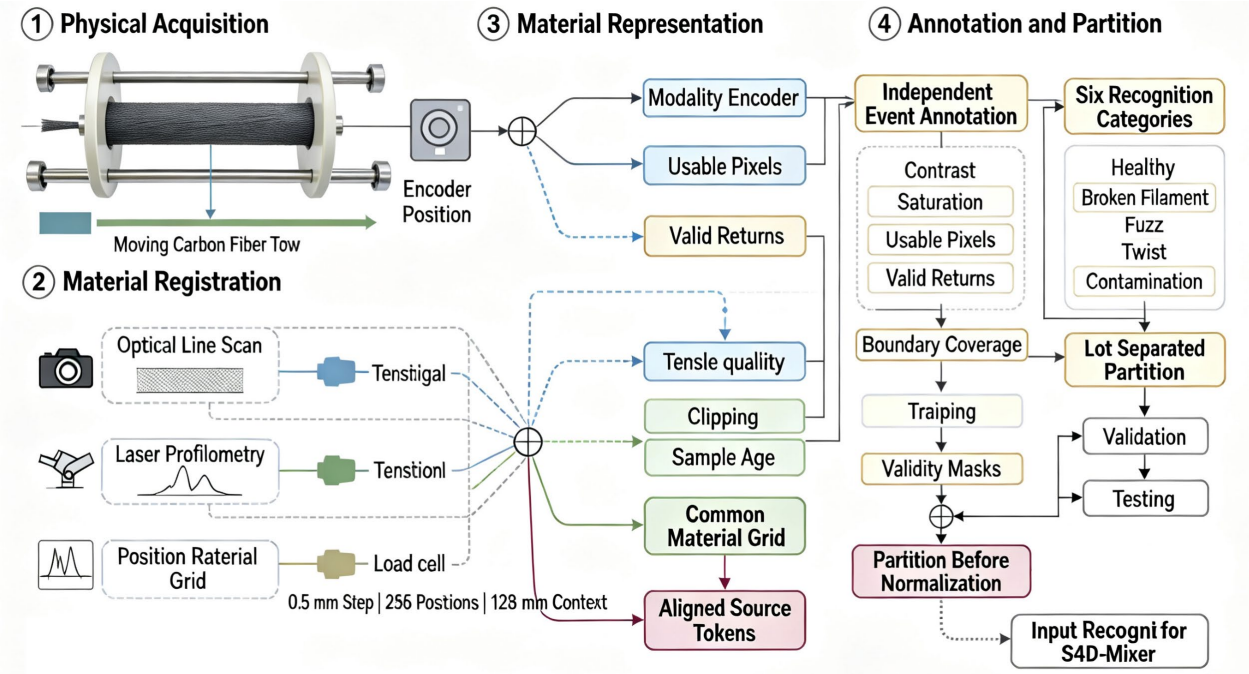


Figure 1. Position aligned acquisition and representation workflow for optical line scans, laser profilometry, and tension sensing

3.3. Annotation Scope and Data Partition Contract

A healthy tow label is assigned only when the observation window shows no visible filament breakage, protruding fuzz, splitting, twist, contamination, or abnormal tension evidence. The label therefore describes the inspected material window and does not imply that later deterioration outside the observation interval has been excluded. Partition sets aside 36 of the 60 lots for training, 12 for validation and 12 for testing, and then generates windows. Continuous events and overlapping material intervals are in the same partition. Step size is 128 and half-window overlap is used. Otherwise, the window counts cannot be used independently, as the manifest will store the source file, lot number, coordinate range and annotation version.

The manifest links each annotation to the corresponding observation and records whether the label has been independently verified. Duplicate events, damaged source files, and missing coordinate records are to be excluded intentionally. The records can audit the revised benchmark without altering the evaluation population secretly. Obtain the data before generating windows, and do not use interpolation to fill in the damaged measurement section. The constructed test scenario assigns 600 windows to each class for transparent comparison. This balanced design makes macro F1, confusion behavior and calibration easier to interpret across classes, but it cannot estimate natural defect prevalence, false alarms per kilometer, or production yield without a length-indexed industrial dataset.

4. S4D Mixer Fusion and Recognition Model

4.1. Stable Causal State Space Processing

Each modality receives a diagonal state space layer whose latent modes have different decay rates and oscillatory components. Transitions are shared within a run, with discretization tied to the spatial grid. State resets at lot boundaries, acquisition discontinuities, and defined long outages. Four residual blocks and 64 modes per channel are initial design settings to be checked through sequence-length and capacity studies.

Within each layer, state h_k accumulates input u_k, while output v_k includes a direct path that preserves short events otherwise attenuated by slow modes. The state-update equation makes the distinction between accumulated context and immediate evidence explicit:

$$h_k = \bar{A}h_{k-1} + \bar{B}u_k, v_k = Ch_k + Du_k \quad (2)$$

The matrices $\bar{A}$ and $\bar{B}$ govern the discrete state transition and input injection, while C projects the latent state into the output representation and D supplies the direct input contribution. The diagonal continuous transition is constrained to have a strictly negative real part. A positive spatial step then produces a discrete transition with magnitude below one for each mode. Complex modes are stored in conjugate pairs so that their combined readout remains real. This constraint

applies to the individual state dynamics; it does not guarantee robustness of the entire nonlinear classifier under arbitrary distribution shift. Stable parameterization and zero-order-hold discretization are coupled through the following equation:

$$a_i = -softplus(\rho_i) + j\omega_i, \bar{A}_i = e^{a_i \Delta x}, \bar{B}_i = \frac{e^{a_i \Delta x} - 1}{a_i} B_i \tag{3}$$

Here, ρ_i controls decay, ω_i controls oscillation, and j is the imaginary unit. The softplus transformation maintains a positive decay magnitude before the negative sign is applied. Diagonal discretization is elementwise. Training can evaluate the recurrence as a causal convolution, while inference updates stored state using the same parameters. Both paths require a numerical equivalence test. Recurrent cost depends on state and channel dimensions; training convolution remains sequence-length dependent. Sensor transfer, registration, and decision handling belong in a separate station latency measurement.

4.2. Reliability Gate and Residual Channel Mixer

At each material position, the fusion gate receives the modality embedding, its quality descriptor, and an availability mask. A small shared scoring network produces a reliability logit for each source. The mask excludes an unavailable source exactly; the remaining scores are normalized over available sources only. When every source is absent, the classifier is bypassed and the output becomes a review request with an acquisition fault indicator. This exception prevents undefined normalization and avoids assigning a confident category to a window containing no observations.

For positions with at least one available source, the gating equation assigns normalized weights while preserving exact exclusion of missing inputs:

$$w_k^{(m)} = \frac{a_k^{(m)} exp\left[g\left(z_k^{(m)}, q_k^{(m)}\right)\right]}{\sum_{r=1}^{M} a_k^{(r)} exp\left[g\left(z_k^{(r)}, q_k^{(r)}\right)\right]} \tag{4}$$

In this equation, $a_k^{(m)}$ denotes binary availability, $z_k^{(m)}$ is the modality embedding, $q_k^{(m)}$ contains measured quality descriptors, and g scores source reliability. The number of sensing sources is M. Gate outputs are diagnostic quantities rather than causal explanations. A low-tension weight can reflect weak discriminatory information or poor measurement quality; separating these interpretations requires quality descriptors and controlled source interventions. Independent modality removal and minimum source exposure reduce collapse onto one stream. Uniform weights are not imposed because the gate must adapt to input quality.

Weighted modality features are concatenated and projected to 192 fused channels. A residual channel mixer applies a two-layer feedforward transformation at each position after temporal processing. Its skip connection preserves the fused representation so that mixing need not overwrite useful modality-specific evidence. Layer normalization operates across channels and does not combine statistics from future positions. The residual block equation separates longitudinal processing from feature interaction:

$$Z' = Z + S4D(LN(Z)), Z'' = Z' + W_2\sigma\left(W_1 LN(Z')\right) \tag{5}$$

The first residual operation introduces causal temporal context, whereas the second transforms channels at the same material coordinate. Here, W_1 and W_2 are learned projections, and σ is the activation function. A symmetric window mixer would introduce future observations despite a causal state branch, so it is excluded from this specification. Readout combines the current representation with a causal context average before producing six logits.

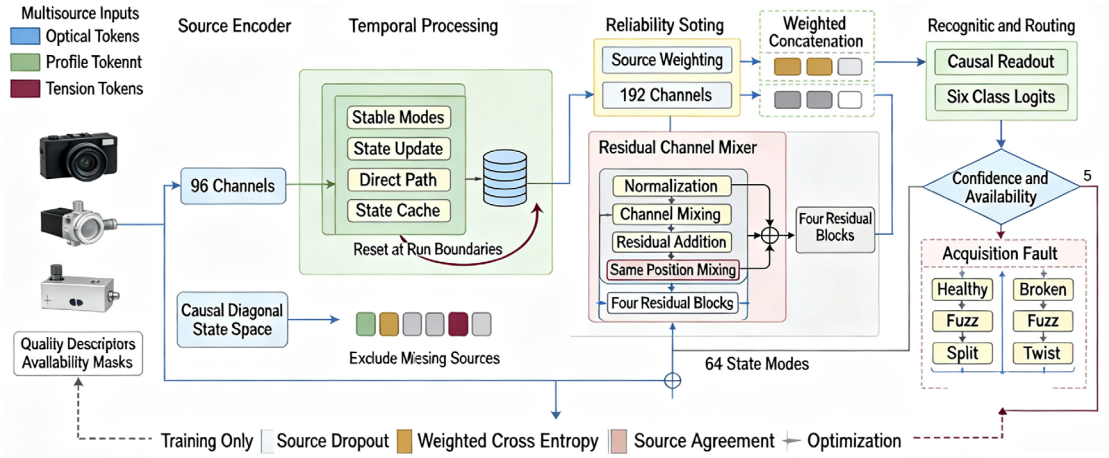


Figure 2. Causal S4D-Mixer with source quality gating, residual channel mixing, calibration, and selective recognition

Figure 2 shows encoders, state updates, reliability weighting, residual mixing, classification, calibration, and review routing, with separate training source-removal and inference state-cache branches.

4.3. Optimization Calibration and Streaming Decisions

Optimization combines class-weighted cross entropy, a modest cross-source agreement term, and weight decay. Class weights are computed from training labels and capped to prevent a rare category from dominating the gradient. The agreement term compares projected features only for sources that are simultaneously valid; an absent source is not required to imitate a present one. This term encourages correspondence at the same material position while retaining the distinct information carried by appearance, geometry, and tension. Excessive agreement could erase useful modality differences, so its coefficient is selected through validation. These objectives enter the training equation as separate contributions:

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda_a \mathcal{L}_{agree} + \lambda_w \|\theta\|_2^2 \tag{6}$$

The classification term averages over labeled windows, the agreement term averages over valid source pairs, and the final term regularizes trainable parameters θ. The coefficients λ_a and λ_w control agreement and regularization, respectively. The proposed schedule uses AdamW, a learning rate of 0.0003, a batch size of 32, and a maximum of 100 epochs. Validation macro F1 determines the checkpoint, while calibration and review thresholds are fitted only after that checkpoint is fixed. Independent modality removal uses a probability of 0.15 per source during training. Windows with no remaining source are omitted from the classification loss and retained for testing the acquisition fault path. This schedule defines an implementable starting point rather than a completed optimization experiment.

Scale the final logits with a positive scalar fitted to the validation predictions for temperature. Only when the calibrated confidence exceeds the upper bound and at least one usable source is available will the prediction be stored. The Threshold for the illustrative deployment analysis is 0.90. The scale of production should cover the costs of defects, additional inspections and spare inspection capacity. Calibration changes the confidence but does not alter the order of class logits; therefore, only by means of temperature scaling can category accuracy be improved. The review policy should also include extended acquisition failures due to operational reasons rather than a low-confidence material decision.

Streaming inference reuses the state to prevent redundant computation of overlapping context. Set run boundary reset state to zero and enable a warm-up flag. Window reconstruction and incremental updates have different time contracts. Report the window latency, per-position update time and station delay; specify the hardware and precision; and include the transfer. Event association is to predict a location in a certain space and retain the original coordinates; near defects can also be added.

5. Experimental Design and Numerical Analysis

5.1. Evaluation Protocol and Recognition Behavior

18,000 windows comprise 10,800 training, 3,600 validation, and 3,600 test windows. Research on grouped validation identifies dependence between related samples as a source of optimistic estimates [17]. The lot partition therefore precedes window generation. The proposed comparison supplies all five models with identical aligned inputs, partition manifests, and checkpoint selection rules. Inspection dataset reporting also identifies acquisition conditions and annotation procedures as factors limiting the transferability of recognition scores [18]. Table 1 lists the example performance and computational indicators. The balanced test set contains 600 windows per category, making category errors directly comparable while leaving production prevalence and inspected material length unspecified.

Table 1. Constructed illustrative recognition and computational comparison

Model	Macro F1 (%)	Average Precision	Latency (*ms*)	Parameters (*M*)
CNN	89.4	0.928	8.6	2.4
TCN	91.2	0.941	11.2	3.1
Transformer	93.6	0.962	24.8	5.8
S4D	94.5	0.973	13.1	3.4
S4D-Mixer	96.7	0.988	15.4	4.2

Class sensitive evaluation distinguishes balanced recognition performance from accuracy dominated by a frequent category [19]. Within the constructed benchmark, Figure 3(a) presents six class F1 distributions across five models and

illustrates how multisource reliability gating would affect rare but operationally important defect states under controlled class balance.

Figure 3(c) shows correct counts of 586, 584, 578, 581, 575, and 576, yielding 3,480 correct decisions and 96.67% accuracy. The error distribution is more informative than this aggregate: eight twist windows are assigned to split and seven split windows to twist, whereas contamination has eight assignments to healthy tow. The first pair calls for inspection of persistent geometric changes; the second concerns missed defects rather than category substitution. Figure 3(d) gives recalls of 97.67%, 97.33%, 96.33%, 96.83%, 95.83%, and 96.00% in category order. Twist exceeds the 95% target by only 0.83 percentage points, compared with 2.67 points for healthy tow. A single common threshold therefore provides unequal category margins even though every displayed recall reaches the target.

The balanced counts expose different costs of error. Healthy tow produces five assignments to fuzz and four to twist, creating unnecessary interventions; eight contamination windows assigned to healthy tow would remain unflagged. Fuzz has eight assignments to broken filament, while broken filament has five assignments to fuzz, motivating examination of attachment and longitudinal extent. These cases require their original coordinate intervals. Gradual transitions may place adjacent windows on opposite sides of an annotation boundary. Independent event records should therefore distinguish boundary disagreement from failure to recognize visible evidence.

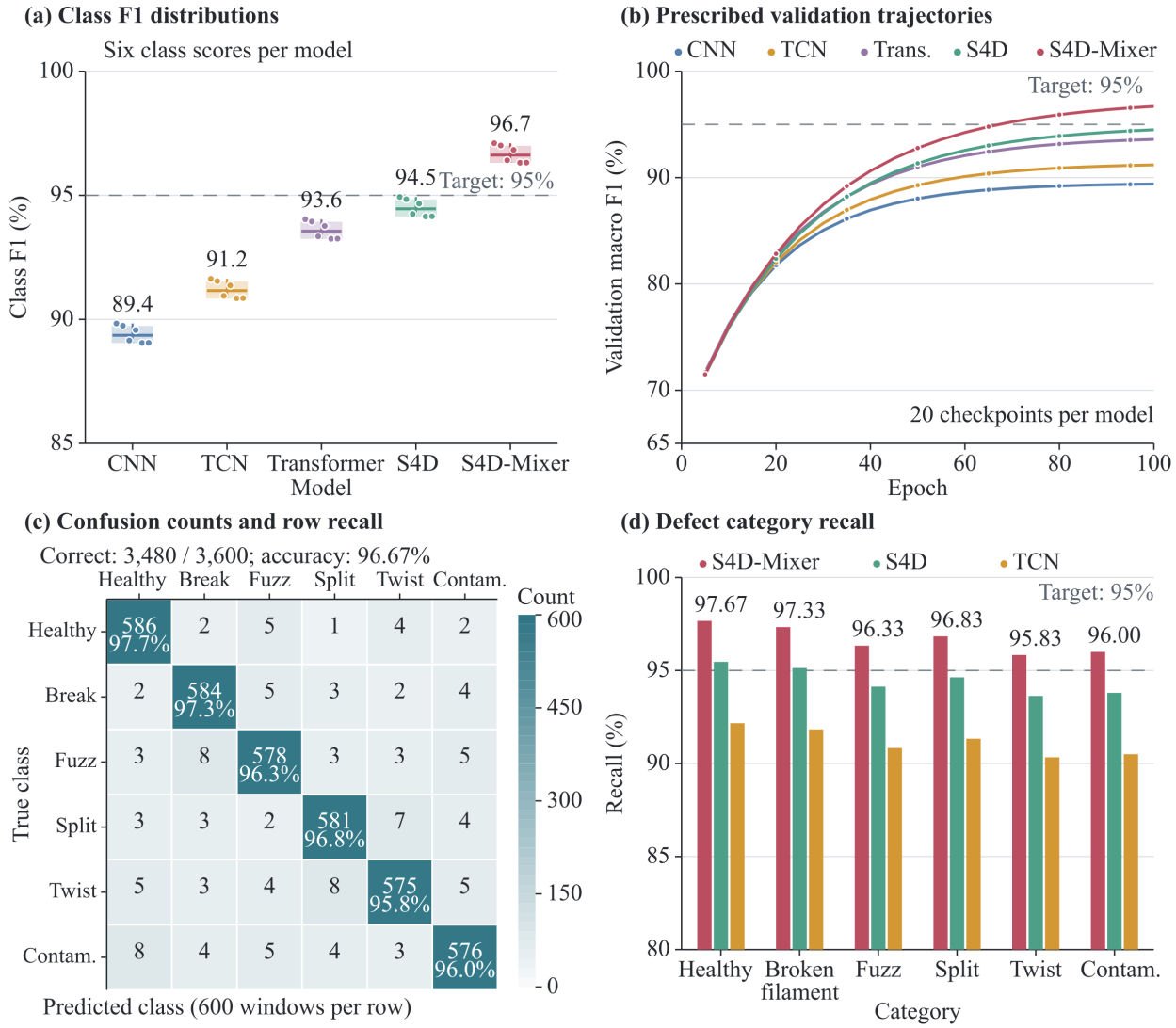


Figure 3. Illustrative recognition comparison: (a) six class F1 distributions across five models; (b) prescribed validation trajectories; (c) six class confusion counts; (d) category recall

Figure 4(a) presents reliability points in ten confidence bins, supported by 120 to 620 windows per bin and 3,600 windows overall. These calibration quantities are computed from the constructed predictions and should be read through the distinction between probability calibration and discrimination [20]. A production study would still need to confirm whether the same confidence margin persists under illumination drift, speed changes, fiber-grade variation and sensor aging, because uncertainty behavior can change under distribution shift [21].

Figure 4(b) compares precision recall curves with illustrative average precision values of 0.928, 0.941, 0.962, 0.973, and 0.988. The difference between the complete model and S4D is 0.015, although the operational significance depends on the recall region selected for inspection. A strong area value does not specify the missed defect rate at one chosen threshold. Figure 4(c) displays five lot scenarios per model, with full ranges of 1.6 to 2.0 percentage points. The complete model remains above the 95% reference across these scenarios, while the S4D baseline straddles it. These ranges describe constructed lot values rather than confidence intervals; a genuine uncertainty estimate would require independently sampled lots and their original predictions.

Figure 4(d) gives selective risks of 0.4%, 0.6%, 0.9%, 1.3%, 2.0%, and 3.3% at coverages from 50% to 100%. Moving from 70% to 80% coverage increases risk by 0.4 percentage points and reduces the review fraction from 30% to 20%. The 1% target is therefore met at the former displayed point and exceeded at the latter. These coordinates do not determine the threshold associated with a particular coverage, because raw confidence values are not supplied. At 90% coverage, an example 2.0% retained error risk applies only to automatically retained windows. Errors and workload among the remaining reviewed windows require separate records, including whether several reviewed windows correspond to one continuous material event.

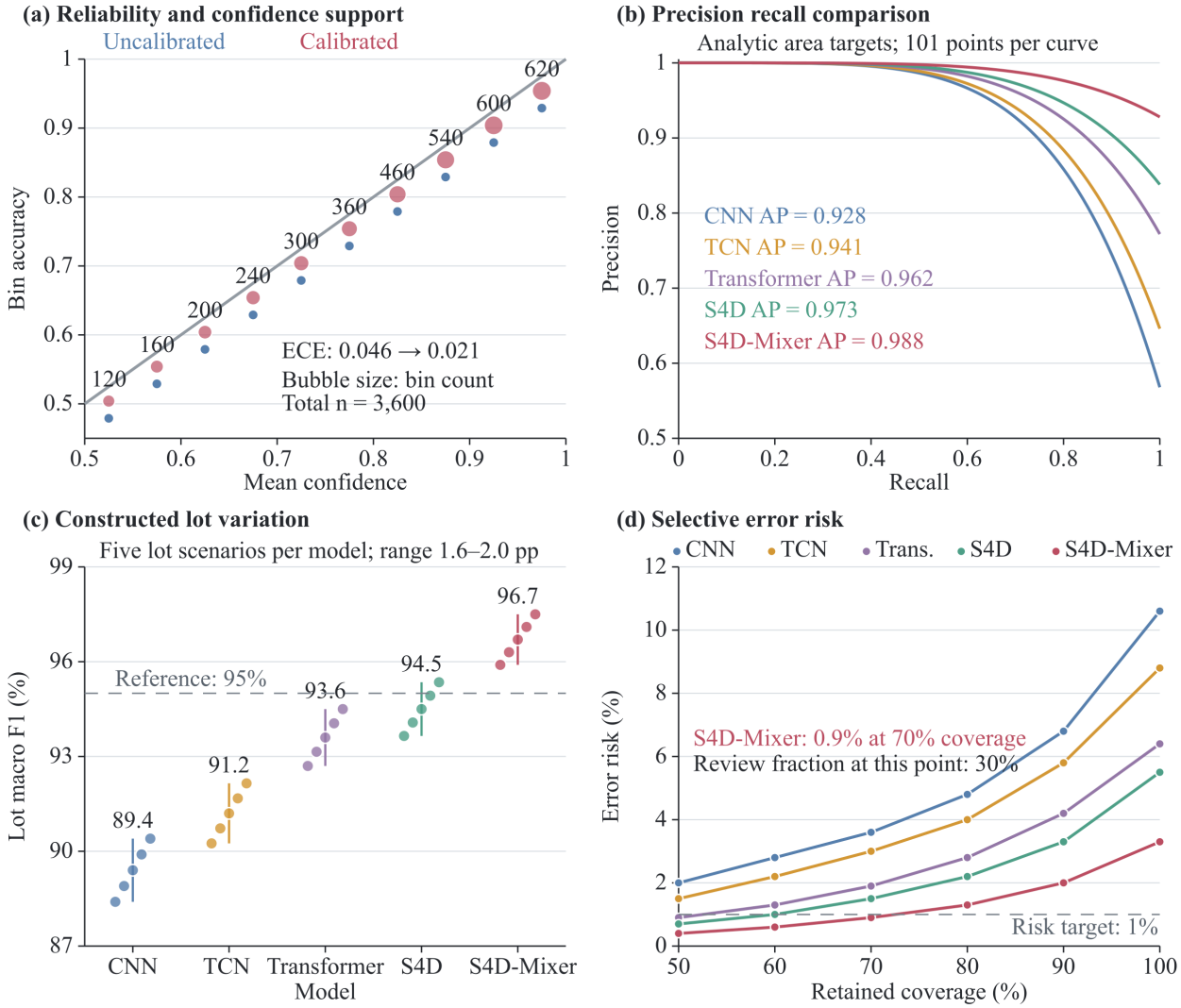


Figure 4. Illustrative prediction quality: (a) reliability and bin support; (b) precision recall curves; (c) constructed lot variation; (d) selective risk and retained coverage

5.2. Source Contribution Ablation and Degradation

Figure 5(a) compares seven source availability patterns evaluated as inference-time sensing conditions. Full sensing gives 96.7% macro F1; optical plus profile, optical plus tension, and profile plus tension describe specified source combinations rather than independently retrained production systems. This framing follows the need to examine missing-modality patterns separately from ordinary random test variation [22].

Controlled ablation methodology requires the removed mechanism and the retained training conditions to be specified [23]. Figure 5(b) gives 93.6% for temporal convolution with concatenation, 94.4% for mixing without S4D, 95.1% for

S4D with fixed fusion, 95.5% without quality descriptors, and 96.7% for the complete model. Removing quality descriptors costs 1.2 percentage points; the complete configuration exceeds fixed fusion by 1.6 points. These prescribed differences motivate separate checks of temporal memory, feature interaction, and quality conditioning, but capacity differences prevent attributing every increment to one mechanism.

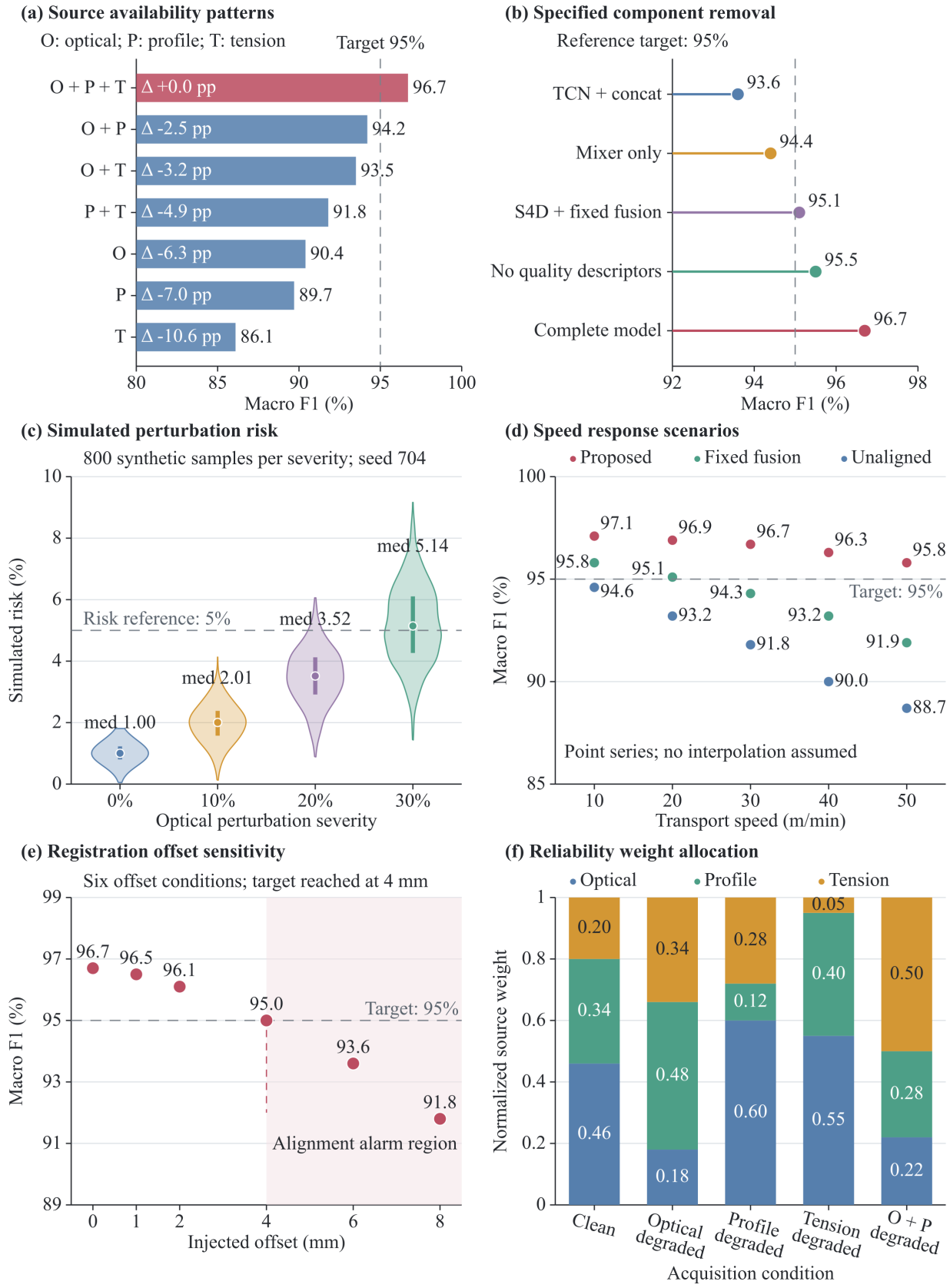


Figure 5. Illustrative source and robustness analyses: (a) availability patterns; (b) component removal; (c) simulated perturbation risk distributions; (d) speed response series; (e) alignment offset sensitivity; (f) reliability weight allocations

Figure 5(c) uses 800 simulated risk samples per optical perturbation severity. Median risks are approximately 1.00%, 2.01%, 3.52%, and 5.14% for severities of 0%, 10%, 20%, and 30%. The severe distribution extends substantially above the 5% reference, while the broadening at intermediate severity indicates that a single mean risk would hide heterogeneous responses. These distributions are simulated examples rather than model prediction uncertainties.

Figure 5(d) shows five speed conditions from 10 to 50 m/min. The complete model declines from 97.1% to 95.8%, fixed fusion from 95.8% to 91.9%, and unaligned concatenation from 94.6% to 88.7%. The respective losses are 1.3, 3.9, and 5.9 percentage points. Research on acquisition shift treats changes in sensing conditions separately from ordinary random test sampling [24]. In an acquisition experiment, speed would also change sampling density and possibly illumination exposure, so these mechanisms would need to be isolated before interpreting the decline. Figure 5(e) gives 96.7%, 96.5%, 96.1%, 95.0%, 93.6%, and 91.8% at alignment offsets of 0, 1, 2, 4, 6, and 8 mm. The score loses 1.7 points by 4 mm and a further 3.2 points between 4 and 8 mm. The illustrated alarm region therefore begins beyond the last displayed offset meeting the recognition target; it is a reporting boundary, not a validated encoder tolerance.

Figure 5(f) examines source weight allocations in five acquisition states. The reported allocations are representative values from constructed scenarios and should be interpreted as an explanation of the gate's intended behavior, not as measured source reliabilities from an installed inspection station.

5.3. Computational Cost Operating Thresholds and Failure Analysis

Figure 6 places five models in the latency versus macro F1 plane, with parameter count encoded by bubble area. The example latencies are 8.6, 11.2, 24.8, 13.1, and 15.4 ms; parameter counts are 2.4, 3.1, 5.8, 3.4, and 4.2 million. Inference benchmarking protocols require batch size, numerical precision, synchronization, and transfer inclusion to be documented [25]. The plotted contract is batch one processing of a complete 256 position window, excluding acquisition and sensor transfer. The complete model lies inside the 95% recognition and 20 ms budget region, with a 4.6 ms timing margin and a 1.7 percentage point recognition margin. Relative to S4D, it adds 2.3 ms and 0.8 million parameters for an example 2.2-point score increment. These increments correspond to approximately 17.6% additional latency and 23.5% additional parameters, exposing a cost that the absolute score alone would conceal. The additional capacity would need to be justified by the defect categories and sensing conditions for which it improves decisions. Relative to the transformer, it reduces latency by 9.4 ms and parameters by 1.6 million. These ratios quantify the illustrated tradeoff without establishing hardware performance or statistical significance.

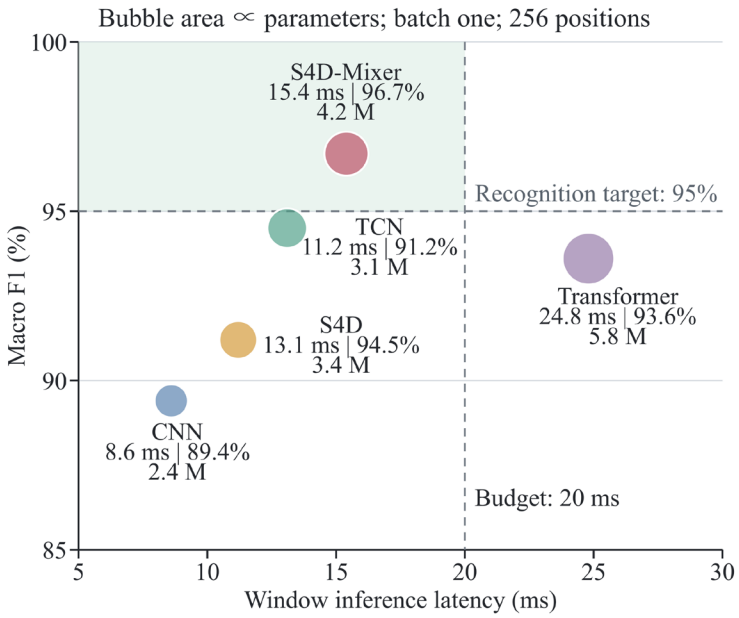


Figure 6. Illustrative deployment tradeoffs

The timing margin must be assessed against processing variability and transport delay. An average latency value alone would not demonstrate that every decision meets a 20 ms in-line deadline, because acquisition buffering, host transfer, reset handling at run boundaries and worst-case scheduling delay must also be measured. The operating threshold should therefore be interpreted together with the risk-coverage behavior of selective recognition [26].

The recognition margin under the sensing conditions is not shown in the illustrative comparison. Practical failure analysis should record low profile return, optical saturation, speed fluctuation and tension spikes at the same material coordinate so that a rejected or incorrect window can be traced to the source condition that caused it.

Link the operational decision to the sources available at the time of failure in the failure log. A relatively large tensile force will also occur at the time of a drop in value, and one needs to determine the reason by adding missing classifications; otherwise, it cannot be assumed that the tension has returned to its original state without such evidence. The log needs to save the quality score, raw source observation, encoder offset, calibrated probability and the length of the degraded interval. If the throughput is lower than the arrival rate, then due to queuing delay, the final intervention will occur after the desired material position, even if the classification is correct. Distinguish the deadline failure from a category error in the record. A retained prediction that uses old profile samples should be marked with the age of the sample to distinguish it from the three simultaneous observations of three sources. Compare the above cases with the truly aligned windows to determine whether the problem is registration or feature fusion. State reset boundaries create another diagnostic scenario: due to the absence of a warm-up after a discontinuity, the previous context that would normally resolve a split or twist has been lost. The preserved decisions near the border should be compared with the decisions obtained from continuous runs in the same class condition. Store the duration of sensor failure, warm-up distance and the separation between the prediction and intervention coordinates as attributes of a single operating event after merging adjacent windows.

6. Conclusion

This paper introduces an S4D-Mixer multisource fusion model for carbon fiber tow defect recognition. The proposed framework aligns optical appearance, laser geometry and tension information on a material-coordinate grid, uses causal diagonal state-space processing to retain sequence context, and applies reliability-aware fusion before selective recognition. The manuscript positions the model as a structured inspection design that can be evaluated under source degradation, calibration and operating-threshold constraints. At present, the study does not include independently obtained production data, executed model comparisons, or hardware timing measurements. These limitations affect claims about transferability, annotation robustness and deployment readiness, and they should be addressed before the model is treated as a validated industrial inspection system.

Future work will collect a versioned dataset of production campaigns, fiber grades, sensing conditions and transport regimes, with independent annotation and lot-separated training, validation and testing partitions. The next validation stage should also measure acquisition-to-decision latency, source failure modes, calibration drift and false alarms per kilometer so that the framework can be assessed under operational manufacturing constraints.

7. Declarations

7.1 Author Contributions

Jindřich Kopecký contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. 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 Date 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 ChatGPT 5.6 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] Wang, H., Luo, H., Zhang, X., Zhao, Z., Wang, J., & Li, Y. (2024). Automatic defect detection of carbon fiber woven fabrics using machine vision. Mechanics of Advanced Materials and Structures, 31(28), 10921-10934. https://doi.org/10.1080/15376494.2023.2299933

[2] Li, Z.-C., Liu, X., Li, M., & Bai, H. (2025). Research on lightweight model of multi-scale target defect detection on tow surface. Manufacturing Automation, 47(6), 114-125. https://doi.org/10.3969/j.issn.1009-0134.2025.06.015

[3] Sun, H., Yang, S., Wang, J., Wu, C., Chen, X., Wang, J., & Chen, K. (2025). Enhanced quantitative detection of delamination defects in large-tow carbon fiber composites using pulsed thermography. Polymer Composites, 46(12), 11164-11176. https://doi.org/10.1002/pc.29677

[4] Zhao, R., Su, Y., Zhang, Y., Wang, J., & Zhou, W. (2025). Defect Detection of Carbon Fiber Cables by Electromagnetic Tomography Based on Proximal Gradient Method. IEEE Transactions on Instrumentation and Measurement, 74, 1-11. https://doi.org/10.1109/tim.2025.3587356

[5] Tytko, G., Rogala, T., Katunin, A., & Yin, W. (2024). Damage Detection in Carbon Fiber-Reinforced Composite Structures Using Eddy Current Pot-Core Sensor. IEEE Access, 12, 123609-123620. https://doi.org/10.1109/access.2024.3454015

[6] Liu, P., Zhang, X., Huang, Y., & Zhang, K. (2026). Defect Characterization and Prediction for Femtosecond Laser Drilling of Carbon Fiber-Reinforced Polymer Based on Image Processing and Machine Learning. Polymer Composites, 47(17), 15414-15427. https://doi.org/10.1002/pc.71295

[7] Gu, M., Zhou, J., Pan, R., & Gao, W. (2025). DF-YOLO: An attempt on enhancing generalization in fabric defect detection based on YOLO network. Textile Research Journal, 95(7-8), 748-766. https://doi.org/10.1177/00405175241269163

[8] Chen, C., Zhou, Q., Li, S., Luo, D., & Tan, G. (2025). Fabric defect detection algorithm based on improved YOLOv8. Textile Research Journal, 95(3-4), 235-251. https://doi.org/10.1177/00405175241261092

[9] Guo, P., Liu, Y., Wu, Y., Wang, J., & Li, Y. (2026). CP-YOLOv8s algorithm for tiny and extreme-type fabric defect detection. Textile Research Journal, 96(19-20), 2005-2024. https://doi.org/10.1177/00405175251362869

[10] Wu, T., & Xie, F. (2026). MSI-YOLO: an efficient model for multiscale fabric defect detection based on improved YOLOv11. Textile Research Journal, 96(13-14), 1444-1463. https://doi.org/10.1177/00405175251329936

[11] Hao, C., Wang, Y., Cui, C., Dong, D., & Mei, D. (2026). Fabric defect detection based on improved YOLOv8s and machine vision. Textile Research Journal, 96(11-12), 1285-1298. https://doi.org/10.1177/00405175251348141

[12] Khanam, R., Hussain, M., Hill, R., & Allen, P. (2024). A Comprehensive Review of Convolutional Neural Networks for Defect Detection in Industrial Applications. IEEE Access, 12, 94250-94295. https://doi.org/10.1109/access.2024.3425166

[13] Deepti Raj, G., & Prabadevi, B. (2024). MoL-YOLOv7: Streamlining Industrial Defect Detection With an Optimized YOLOv7 Approach. IEEE Access, 12, 117090-117101. https://doi.org/10.1109/access.2024.3447035

[14] Liu, Y., & Gao, Y. (2025). Multi-Modal Dynamic Fusion for Defect Detection in Electronic Products: A Novel Approach Based on Energy and Deep Learning. IEEE Access, 13, 118565-118573. https://doi.org/10.1109/access.2025.3584551

[15] Shao, Z., Wang, H., Cai, Y., Chen, L., & Li, Y. (2025). UA-Fusion: Uncertainty-Aware Multimodal Data Fusion Framework for 3-D Object Detection of Autonomous Vehicles. IEEE Transactions on Instrumentation and Measurement, 74, 1-16. https://doi.org/10.1109/tim.2025.3548184

[16] Tellamekala, M. K., Amiriparian, S., Schuller, B. W., André, E., Giesbrecht, T., & Valstar, M. (2024). COLD Fusion: Calibrated and Ordinal Latent Distribution Fusion for Uncertainty-Aware Multimodal Emotion Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(2), 805-822. https://doi.org/10.1109/tpami.2023.3325770

[17] Wang, J., Li, J., Wang, R., & Zhou, X. (2024). VAE-Driven Multimodal Fusion for Early Cardiac Disease Detection. IEEE Access, 12, 90535-90551. https://doi.org/10.1109/access.2024.3420444

[18] Sreevidya, P., Aravinth, J., & Samiappan, S. (2024). Intermediate Layer Attention Mechanism for Multimodal Fusion in Personality and Affect Computing. IEEE Access, 12, 112776-112793. https://doi.org/10.1109/access.2024.3442377

[19] Argha, A., Alinejad-Rokny, H., Baumgartner, M., Schreier, G., Celler, B. G., Redmond, S. J., Butcher, K., Ooi, S.-Y., & Lovell, N. H. (2025). A Novel Deep Ensemble Method for Selective Classification of Electrocardiograms. IEEE Transactions on Biomedical Engineering, 72(2), 833-842. https://doi.org/10.1109/tbme.2024.3476088

[20] Balanya, S. A., Maroñas, J., & Ramos, D. (2024). Adaptive temperature scaling for Robust calibration of deep neural networks.

Neural Computing and Applications, 36(14), 8073-8095. https://doi.org/10.1007/s00521-024-09505-4

[21] Tiezzi, M., Casoni, M., Betti, A., Guidi, T., Gori, M., & Melacci, S. (2025). Back to recurrent processing at the crossroad of transformers and state-space models. Nature Machine Intelligence, 7(5), 678-688. https://doi.org/10.1038/s42256-025-01034-6

[22] Tang, Q., Dai, C., Wu, Y., & Zhou, H. (2024). MLP-Mixer based Masked Autoencoders are Effective, Explainable and Robust for Time Series Anomaly Detection. Proceedings of the VLDB Endowment, 18(3), 798-811. https://doi.org/10.14778/3712221.3712243

[23] Olalde-Verano, J. I., Kirch, S., Pérez-Molina, C., & Martín, S. (2025). SambaMixer: State of Health Prediction of Li-Ion Batteries Using Mamba State Space Models. IEEE Access, 13, 2313-2327. https://doi.org/10.1109/access.2024.3524321

[24] Chen, J., Tan, X., Rahardja, S., Yang, J., & Rahardja, S. (2024). Joint Selective State Space Model and Detrending for Robust Time Series Anomaly Detection. IEEE Signal Processing Letters, 31, 2050-2054. https://doi.org/10.1109/lsp.2024.3438078

[25] Yang, M., & Chen, L. (2025). HC-Mamba: Remote Sensing Image Classification via Hybrid Cross-Activation State-Space Model. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18, 10429-10441. https://doi.org/10.1109/jstars.2025.3555217

[26] Feng, H., Chakravarty, S., & Lonardi, S. (2025). Prediction of DNA Methylation With Long-Range State-Space Models. IEEE Transactions on Computational Biology and Bioinformatics, 22(6), 3584-3590. https://doi.org/10.1109/tcbbio.2025.3634738