

Article

Data-Driven Physics-Informed Fusion for Clothing Material Identification in Washing Machines

Shurong Zhang ^{1,†}, Yuze Gao ^{2,†}, Yongtao Wan ², Bin Zhang ^{1,*}  and Jianxiong Zhu ^{2,*} ¹ Xiaomi Technology (Wuhan) Co., Ltd., Wuhan 430010, China; zhangshurong@xiaomi.com² School of Mechanical Engineering, Southeast University, Nanjing 211189, China; 220240467@seu.edu.cn (Y.G.); 220240418@seu.edu.cn (Y.W.)

* Correspondence: zhangbin14@xiaomi.com (B.Z.); mezhujx@seu.edu.cn (J.Z.)

† These authors contributed equally to this work.

Abstract

To meet the demand for refined laundry care in intelligent washing machines and address the low accuracy, poor robustness, and lack of physical interpretability of existing material recognition technologies, a recognition method integrating physical prior knowledge is proposed. Based on a physical experimental platform for drum washing machines, mechanical vibration signals from a three-axis acceleration sensor and motor electromagnetic signals are collected synchronously, a dataset consisting of soft and hard loads is constructed, and time-domain alignment of heterogeneous signals is realized using adaptive pooling technology. Combined with the mechatronic coupling mechanism in the loosening, deviation detection, and weighing stages of washing machines, a Physics-Aware Dual-Stream Multi-Scale Temporal Convolutional Network (PSA-DSMS-TCN) is designed. The network extracts mechanical and electromagnetic features in parallel through a dual-stream structure, expands the receptive field using multi-scale dilated convolution, and introduces an operating condition-gated attention mechanism to achieve dynamic feature fusion. The results of 5-fold cross-validation show that the model achieves an average recognition accuracy of 94.05%, with consistent performance enhancement and substantial practical robustness. The results demonstrate that the PSA-DSMS-TCN effectively improves the precision of material prediction while maintaining lightweight characteristics, providing reliable technical support for the intelligent matching of laundry care parameters.

Keywords: material recognition; Temporal Convolutional Network; physical prior knowledge; dual-stream; washing machine



Academic Editor: Juan Gabriel Avina-Cervantes

Received: 19 January 2026

Revised: 2 March 2026

Accepted: 4 March 2026

Published: 8 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

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

1. Introduction

With the continuous upgrading of consumer expectations, the functional paradigm of modern washing machines has shifted from basic stain removal to comprehensive and refined fabric care. Consumers increasingly demand advanced health-oriented features (e.g., sterilization and mite removal) alongside meticulous fabric protection to minimize mechanical wear and irreversible fiber damage during washing and drying cycles. Given the highly diverse characteristics of modern textiles—ranging from cotton and synthetic fibers to delicate materials such as silk, wool, and down—the requirements for washing algorithms, thermal control, and mechanical agitation vary significantly. Improper parameter matching is a primary cause of irreversible textile degradation. For instance, down garments require specialized tumbling patterns to maintain fluffiness, while protein-based

fibers like silk and fur heavily rely on low-temperature drying or waterless “air-wash” technologies to prevent structural destruction. Consequently, the precise pre-identification of clothing materials has emerged as a critical technological bottleneck in the appliance industry and the fundamental prerequisite for intelligently matching core laundry and drying parameters. However, current clothing material recognition technologies in the washing machine industry suffer from shortcomings such as low accuracy and large errors, making it difficult to meet the care needs of high-end fabrics [1]. Therefore, conducting research on efficient and accurate recognition methods holds significant practical significance for improving the intelligence level of washing machines and optimizing user experience [2].

Existing clothing material recognition technologies have evolved into multiple technical routes, among which mainstream solutions mainly include radio frequency identification (RFID) and computer vision (AI camera). However, all these schemes face significant limitations in practical engineering applications. Specifically, although RFID technology achieves high recognition accuracy, it heavily relies on electronic tags pre-installed when textiles are manufactured, resulting in extremely poor industrial universality and difficulty in covering the massive number of existing clothes owned by consumers. Visual schemes based on AI cameras require the additional deployment of high-cost image sensors, and the water vapor, foam, and low-light environment inside the washing machine drum severely degrade image quality, restricting their robustness. Material identification methods based on fabric water absorption require tedious water intake and standing tests, which not only waste water resources but also substantially prolong the laundry cycle. In contrast, mechanical signal schemes based on acceleration sensors have attracted extensive attention due to their low hardware cost. Nevertheless, existing methods are highly vulnerable to mechanical resonance of the whole machine and environmental noise under complex working conditions, and their recognition accuracy and stability can hardly meet the stringent standards of high-end intelligent laundry care.

With the continuous iteration and in-depth penetration of AI technology, its application in the field of smart home appliances has become increasingly widespread, endowing traditional household appliances with more powerful intelligent functions [3–6]. Nowadays, a growing number of enterprises are deeply integrating AI technology into the R&D and upgrading of washing machines. Through algorithm optimization, data mining, and other means, they continuously improve the intelligence level of products, driving washing machines toward the direction of precise control and intelligent adaptation [7–9]. Existing studies have shown that the introduction of multi-dimensional physical signals such as thermal imaging, combined with hybrid deep learning models, achieves high accuracy and repeatability in material classification [10].

However, purely data-driven models relying only on washing machine sensor data face limitations when dealing with such complex mechatronic systems. On the one hand, due to the highly dynamic nature of load distribution and time-varying characteristics of motor speed during the washing process, traditional “black-box” deep learning models are prone to fitting environmental noise or spurious correlations, resulting in restricted generalization performance under complex and variable operating conditions. On the other hand, the feature extraction process without physical constraints is usually accompanied by poor interpretability, which can hardly meet the stringent requirements of industrial control systems for operational reliability. Therefore, it is particularly critical to incorporate physical prior knowledge into the deep learning framework. By explicitly embedding the inherent physical laws of washing machine operation into the neural network architecture, the hypothesis space of the model can be effectively constrained. This learning paradigm, guided by physical mechanisms, drives the model to adaptively focus on the most mechan-

ically consistent and discriminative mechatronic features, thereby breaking through the bottlenecks of purely data-driven methods in terms of robustness and interpretability.

In practical industrial contexts, this challenge is further exacerbated by the synchronization mismatch between multi-rate sensor signals. Due to the constraints of industrial hardware, mechanical vibrations and electromagnetic signals exhibit significant discrepancies in sampling frequencies and time delays, inevitably leading to information misalignment and feature distortion during fusion. Traditional alignment paradigms, such as interpolation or cross-correlation, are often inadequate; they frequently introduce artifacts or phase errors that destroy critical high-frequency mechanical textures. Consequently, there is an urgent need for a robust fusion framework capable of preserving the raw physical characteristics of asynchronous signals while enabling the network to adaptively extract high-level mechatronic coupling features across these heterogeneous temporal scales without loss of information.

Inspired by multi-physical signal fusion mechanisms, this study proposes an accurate material recognition scheme that deeply integrates physical priors based on the existing hardware resources of washing machines. This scheme deeply explores the interactive features between conventional acceleration sensors and motor parameters, and realizes efficient recognition using deep networks. Without increasing hardware cost or washing time, it effectively improves recognition accuracy and provides a new technical path for personalized laundry care.

2. Results and Discussion

2.1. Design of Clothing Material Recognition Method

To address the issues of non-stationary time-series signals in traditional deep learning models for clothing recognition tasks in washing machines and the lack of physical interpretability in pure data-driven models, this chapter proposes a deep learning architecture integrated with physical prior knowledge—the Physics-State Aware Dual-Stream Multi-Scale Temporal Convolutional Network (PSA-DSMS-TCN)—specifically designed to distinguish between the soft and stiff fabrics illustrated in Figure 1a,b.

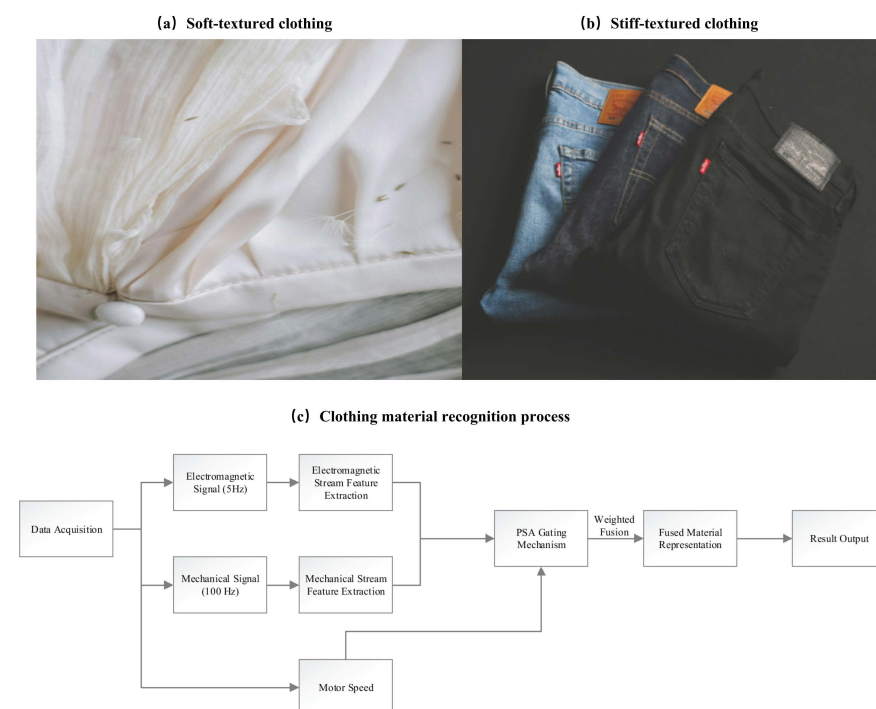


Figure 1. (a) Soft-textured clothing. (b) Stiff-textured clothing. (c) Clothing material recognition process.

The overall architecture is illustrated in Figure 1c. Firstly, an asynchronous data input strategy that preserves the original physical characteristics is adopted for the multi-frequency sampling characteristics of heterogeneous sensors. Subsequently, combining the unique rotational speed step characteristics of washing machines, a dual-stream multi-scale feature extraction network with mechanical-electrical decoupling is designed to capture the long and short temporal dependencies of mechanical vibration and electromagnetic response in parallel. Finally, this study innovatively introduces a rotational speed-based physical state operating condition-gated attention mechanism, explicitly embedding prior operating condition knowledge into the network forward propagation to guide the model to adaptively focus on key physical streams at different stages, thereby learning clothing material representations that conform to mechanical laws [11]. This chapter will elaborate on the theoretical basis and implementation details of each module.

2.2. Heterogeneous Data Preprocessing and Asynchronous Strategy

In the experimental acquisition system constructed in this study, the experimental subject is a Xiaomi washing machine. Restricted by the underlying hardware architecture of industrial-grade equipment, the sampling rate of the three-axis acceleration sensor is 100 Hz, while the sampling rate of motor electromagnetic signals (current, rotational speed, etc.) acquired from the variable-frequency drive board is 5 Hz, as shown in Figure 2a,b. To retain the authentic physical states captured by the sensors to the greatest extent, this study preserves the non-stationary variable-length characteristics and heterogeneous sampling rates of mechanical vibration signals and motor electromagnetic signals. The 100 Hz high-frequency three-axis acceleration data sequence and the 5 Hz low-frequency four-channel electromagnetic data sequence are treated as two completely independent information streams. When feeding data into the deep learning network, no destructive zero-padding or truncation is performed. Instead, the two unequal-length tensors are directly and separately fed into the mechanical branch and electromagnetic branch of the subsequently designed dual-stream network. This native asynchronous decoupling strategy avoids phase errors and high-frequency texture loss caused by forced time-domain alignment.

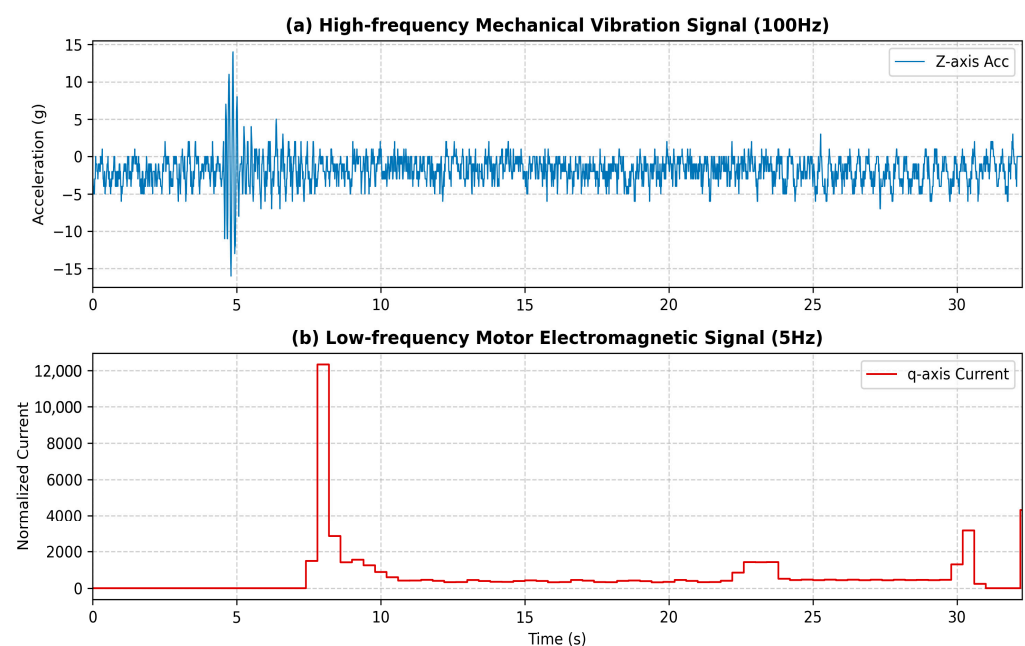


Figure 2. (a) High-frequency mechanical vibration signal. (b) Low-frequency motor electromagnetic signal.

Since the physical dimensions of signal sources such as mechanical vibration, motor current, and rotational speed differ significantly, their value ranges are highly unbalanced. Direct input into the network would lead to numerical instability during gradient descent and cause the model to be overly biased toward feature channels with larger absolute values. To eliminate the dimensional barrier across multimodal data and accelerate model convergence, this paper introduces a channel-wise Z-Score normalization strategy before asynchronous input. For a given feature matrix, the mean and standard deviation of each physical channel within valid time steps are computed, followed by a linear transformation.

$$x'_{i,c} = \frac{x_{i,c} - \mu_c}{\sigma_c} \quad (1)$$

where $x_{i,c}$ denotes the sampled physical signal at the i -th time step of the c -th channel in the original feature sequence. This normalization operation maps signals from all heterogeneous sensors uniformly into a standard normal distribution space with a mean of 0 and a variance of 1, while fully preserving the original waveform oscillation patterns and geometric topological structure. It thus provides a high-quality data foundation for the dual-stream network to extract unbiased mechatronic coupling features.

2.3. Physics-State Aware Dual-Stream Multi-Scale TCN Architecture

Before designing the deep learning model, it is crucial to analyze the physical operation process of the washing machine. Its weighing motion is not uniformly continuous but divided into three steady-state stages with clear physical meanings—loosening, unbalance detection, and weighing—according to the control logic, where the vibration and current signals in different stages contain differentiated physical information about clothing material, distribution and weight respectively. To address this issue, this paper proposes the PSA-DSMS-TCN architecture, which achieves targeted optimization through an explicit physics-stream decoupling and operation-condition gating mechanism.

Traditional RNN/LSTM models suffer from low serial computing efficiency and gradient vanishing when processing long sequences and are also poor at capturing abrupt signals [12]. To extract discriminative features from the high-frequency, noisy motor time-series signals, an improved Temporal Convolutional Network (TCN) is constructed in this study [13,14]. Compared with traditional Recurrent Neural Networks (RNN/LSTM), TCN boasts advantages such as strong parallel computing capability, stable gradient propagation, and flexible and controllable receptive field [15,16].

The core advantage of TCN lies in expanding the receptive field through dilated convolution, which enables the capture of long-range temporal dependencies without recurrent structures. To capture long-range historical dependency information without increasing network depth and the number of parameters, this paper introduces dilated convolution technology [17–19]. For an input sequence x and a convolution kernel f , the convolution operation with a dilation factor of d is defined as:

$$(x *_d f)(s) = \sum_{k=0}^{K-1} f(k)x(s - dk) \quad (2)$$

where $K = 3$ denotes the kernel size. The TCN architecture designed in this paper consists of two parallel feature extraction streams, each composed of 2 stacked residual blocks. At each level, the dilation factor grows exponentially in depth, i.e., $dl = 2^{l-1}$ (yielding 1, 2, and 4 respectively), as illustrated in Figure 3.

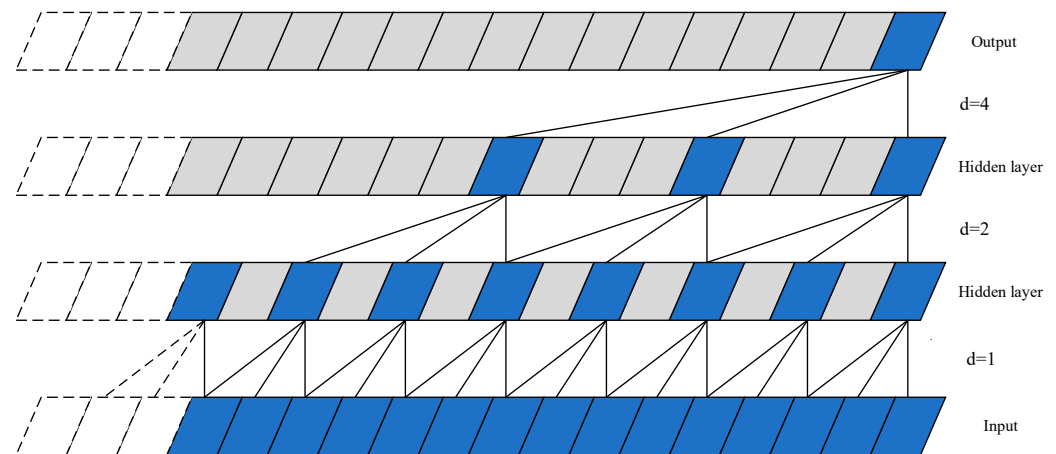


Figure 3. TCN dilated causal convolution structure.

The TCN module is composed of multiple stacked residual blocks. Each residual block contains two dilated convolutional layers, a weight normalization layer, and a ReLU activation function, as illustrated in Figure 4. For lightweight deployment, the base channel width is set to 16 in this paper. The mechanical stream branch gradually upscales the 3-channel input to 32 dimensions via residual blocks; the electromagnetic stream branch similarly maps the 4-channel input to 32 dimensions. To prevent overfitting in small-sample training, a high-rate Dropout ($p = 0.5$) is introduced after the convolutional layers. Furthermore, the introduction of residual connections enables the network to learn identity mappings, effectively alleviating the degradation problem of deep networks. This module gradually maps the original 7-channel physical signals to a high-dimensional feature space, and finally outputs the feature tensor $X_{\text{tcn}} \in \mathbb{R}^{B \times \text{Cout} \times L}$. The formula for residual connections is as follows:

$$o = \text{Activation}(x + F(x)) \quad (3)$$

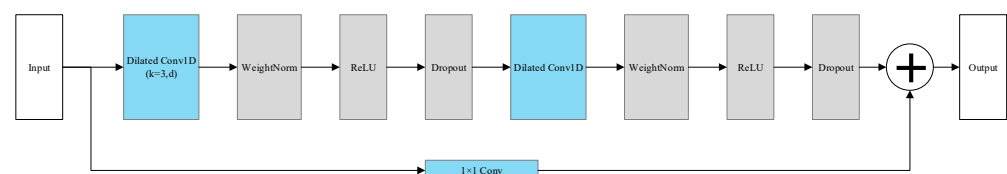


Figure 4. Residual block of TCN.

Considering the significant differences in frequency-domain characteristics and physical representations between mechanical vibration signals and motor electromagnetic signals, direct channel concatenation often leads to information annihilation [20]. Therefore, this paper designs a parallel dual-stream input structure: the input of the mechanical stream includes A_{cc_x} , A_{cc_y} , A_{cc_z} , and rotational speed signals; the input of the electromagnetic stream includes I_d , I_q , I_{u_rms} , and Power [21,22].

To address the inconsistent sampling rates of the two signal streams, the dual-stream network processes them in parallel through independent feature extraction paths. Before feature fusion, a Global Adaptive Average Pooling layer is adopted to achieve automatic alignment in the temporal dimension. Statistics show that the total number of parameters of this model is approximately 35 k, indicating extremely high computational efficiency.

Given the diversity of time scales in the internal features of a single physical stream, a convolution kernel with a single receptive field cannot capture them simultaneously. This paper introduces multi-scale dilated convolutions into the residual blocks of the TCN [23–25], as illustrated in Figure 5. Each feature extraction layer comprises three

parallel convolution branches with dilation rates set to $d = 1, 2, 4$ respectively. The branch with $d = 1$ focuses on local details such as glitches and spikes in acceleration signals; the branches with $d = 2, 4$ greatly expand the receptive field and focus on the long-term dependencies and trend evolution of signals.

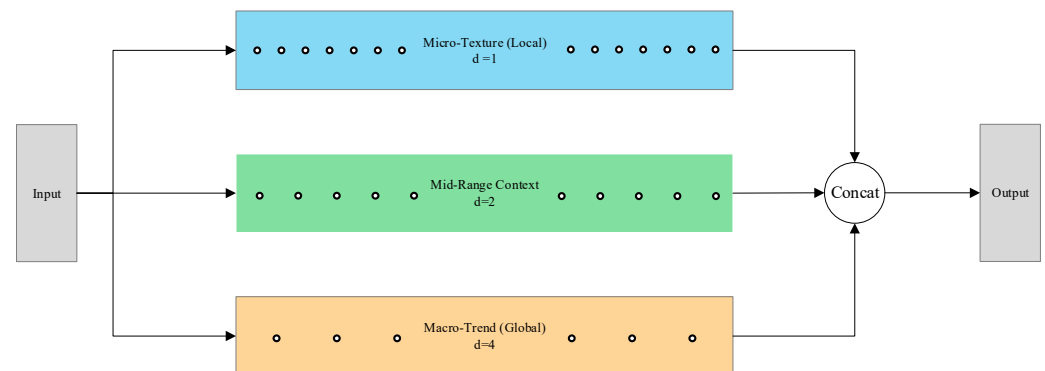


Figure 5. Multi-scale dilated convolution module.

To incorporate physical priors into the model, a lightweight physics-state gating attention module is proposed in this study [26–28], which is shown in Figure 6a. The core idea of this module is to use the rotational speed signal as the gating signal.

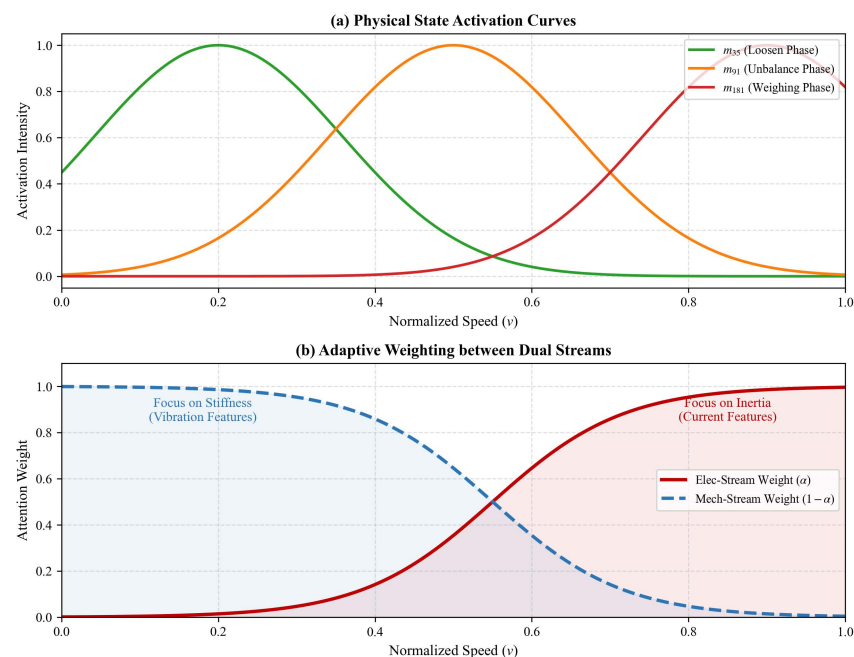


Figure 6. (a) Physical state activation curves. (b) Adaptive weighting between dual streams.

The gating module first extracts the real-time normalized rotation speed signal from the motor stream and maps it into an operating condition activation vector $St = [m35, m91, m181]$ representing the three stages of “loosening—deviation detection—weighing” using three sets of Gaussian kernel functions with clear physical meanings. The smooth transition between operating conditions is described in the form of probability distributions. This design ensures that the feature fusion weights evolve steadily during dynamic rotation speed switching, effectively avoiding oscillations in recognition results caused by signal jitter, while mathematically guaranteeing the global differentiability of the network. The state vector St is fed into a lightweight multi-layer perceptron (MLP). After nonlinear transformation, adaptive weights α along the channel dimension are generated via the

Sigmoid activation function. These weights α are directly applied to the complementary dual-stream feature fusion strategy. To explicitly represent the dominant roles of mechanical and electrical features at different stages, the fused feature tensor is computed as follows:

$$f_{fused} = \alpha f_{elec} + (1 - \alpha) f_{mech} \quad (4)$$

Under this mechanism, when the rotation speed is in the high-speed weighing stage, the weight increases to guide the model to focus on electromagnetic features that reflect load inertia. In the low-speed loosening stage, the weight shifts toward the mechanical stream to capture more discriminative vibration texture features, as illustrated in Figure 6b. With an extremely small number of parameters, this design achieves the explicit embedding of physical prior knowledge into the network structure, significantly enhancing model interpretability.

2.4. Dataset Construction

In this study, a high-fidelity mechatronic signal acquisition platform was built around an industrial-grade Xiaomi front-loading washing machine with typical mechatronic characteristics. The platform integrates a three-axis accelerometer (sampling rate 100 Hz) and a motor variable-frequency drive circuit (sampling rate 5 Hz). An adaptive pooling mechanism is adopted to resolve the sampling step mismatch between heterogeneous sensors, ensuring the spatiotemporal consistency of mechanical vibration and electromagnetic response at the feature level.

The experimental design focuses on the dry state perception stage before water intake. During this critical physical window, the mechanical interactions between fabrics, the inner drum wall, and adjacent fabrics follow pure tribological and multi-body dynamics laws, avoiding the nonlinear masking of signal features by water damping. This enables the capture of highly discriminative physical features that reflect the intrinsic properties of fabrics, such as stiffness, friction coefficient, and inertia.

For sample classification, in collaboration with enterprise performance experts, fabrics were scientifically divided into soft load (including pure cotton fabrics, towels, etc., totaling 92 cycles) and hard load (including high-grammage denim, thick canvas, etc., totaling 76 cycles) according to their mechanical behavior and mass distribution under typical washing cycles. Experimental samples cover the common load range from 1.0 kg to 5.0 kg. A total of 168 independent washing cycles fully include three typical operating conditions in the initial startup stage: loosening, unbalance detection, and weighing.

To ensure reliable validation, experiments strictly follow the principle of physical isolation and cycle decoupling: each sampling corresponds to a complete shutdown–cold start–system initialization procedure. Between consecutive acquisitions, clothes are manually stirred and randomly rearranged to simulate the random eccentric state in real household scenarios, thereby achieving full cycle-level separation in 5-fold cross-validation and fundamentally eliminating the common data leakage problem in time-series data.

To address complex disturbances in industrial environments, a physics-driven data augmentation strategy was introduced. In addition to traditional 10-fold sample expansion, the dataset was extended to 1680 samples by adding physically consistent Gaussian white noise (simulating loose sensor installation and electromagnetic interference) and time-scale perturbation (simulating slight variations in motor response caused by ambient temperature and voltage fluctuations). To prevent data leakage during model evaluation, a cycle-based cross-validation protocol was adopted. The initial 168 independent cycles were partitioned into five folds prior to any data manipulation. Data augmentation was subsequently executed only on the training subsets within each fold, ensuring that the test set remained entirely independent and consisted solely of original samples.

2.5. Experimental Platform and Hardware Deployment Metrics

All model training and inference tasks in this study were completed on a high-performance computing platform. The experimental platform was built on a Lenovo Legion Y7000 (2020) laptop, manufactured by Lenovo Group Limited, Beijing, China, equipped with a 10th-Gen Intel Core i5-10300H processor (2.50 GHz), 16 GB DDR4 2933 MHz system memory, and an NVIDIA GeForce GTX 1650 discrete graphics card (4 GB GDDR6). The software environment was based on the Windows 11 operating system, with Python 3.12 as the programming language and the CUDA 12.4 computing library configured to support the efficient operation of deep learning frameworks.

To verify the deployment feasibility of the PSA-DSMS-TCN architecture in resource-constrained embedded environments, this paper quantitatively evaluates its computational complexity and inference latency: (1) The model adopts a lightweight dual-stream structure (channel width $C = 16$) with a total parameter count of only 35.2 k. At FP32 precision, the model weights occupy only approximately 140.8 KB of storage space, which is significantly lower than the resource thresholds of mainstream industrial-grade control chips (e.g., STM32 series), demonstrating distinct advantages for hardware deployment. (2) Simulated tests are conducted by constructing a single-batch mechatronic asynchronous signal stream (3200 points for the mechanical stream and 160 points for the motor stream). The results show that the average inference latency of the model is only 3.13 ms. Considering that the operating step size of industrial washing machine control systems is typically 100–200 ms, the computational overhead of this model accounts for only 1.5~3.1% of the system task cycle. This indicates that while maintaining high accuracy, the architecture features extremely low computational cost and real-time response capability, fully meeting the online recognition requirements of variable-frequency drive boards with low computing power.

2.6. Experimental Results

To comprehensively and objectively evaluate the performance of the AI model in the task of laundry material recognition, this study selected accuracy, precision, recall, and the F1-score as the core quantitative metrics. Meanwhile, the confusion matrix was introduced as a visual analysis tool to reveal the detailed misclassification of the model across different categories.

To eliminate the impact of randomness in dataset division on the evaluation results, this study adopted a rigorous 5-fold cross-validation method, and the statistical experimental results are shown in Table 1 and Figure 7a.

Table 1. Model performance statistics based on 5-fold cross-validation.

Fold	Accuracy	Precision	Recall	F1-Score
Fold 1	97.06%	96.88%	97.37%	0.9704
Fold 2	97.06%	97.5%	96.67%	0.9699
Fold 3	88.24%	88.93%	87.85%	0.8807
Fold 4	90.91%	91.67%	91.67%	0.9091
Fold 5	96.97%	97.37%	96.67%	0.9692
Mean	94.05%	94.47%	94.05%	0.9399

Statistical data demonstrate that the PSA-DSMS-TCN model exhibits excellent classification performance and robustness: it achieves an average precision of 94.47%, which can effectively prevent soft loads from being misclassified as stiff loads and eliminate the risk of damaging valuable clothes due to excessive mechanical force; its average recall rate reaches 94.05%, enabling comprehensive detection of stiff loads and ensuring the washing ratio of heavily soiled clothes meets the standard.

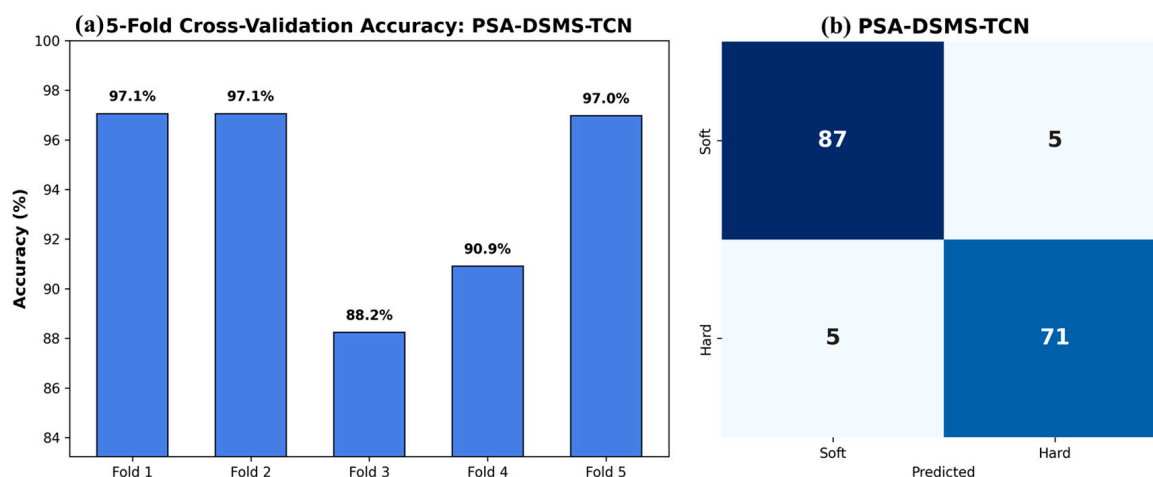


Figure 7. (a) Accuracy comparison of PSA-DSMS-TCN model with 5-fold cross-validation. (b) Confusion matrix of the PSA-DSMS-TCN model.

It is worth noting that a slight performance drop was observed in Fold 3 during cross-validation. This fluctuation is likely attributed to the stochastic nature of fabric distribution inside the drum. Specifically, certain cycles in Fold 3 exhibited quasi-balanced states for stiff fabrics, where the mechatronic signatures temporarily resembled those of softer loads. Such overlap in physical characteristics presents a common challenge for data-driven models. Nevertheless, the robust average performance (94.05% accuracy) confirms that the proposed PSA-DSMS-TCN maintains strong generalization capabilities despite the inherent non-stationarity of real-world operating conditions.

Figure 7b presents the confusion matrix of the model on the test set. It can be observed that the color of the main diagonal area is significantly deepened with concentrated values, which intuitively reflects a high degree of consistency between the model's predicted results and the true labels. Specifically, the model has established a clear classification boundary between soft and hard materials, achieving accurate differentiation of the two core material categories with an extremely low misclassification rate.

To verify the contribution of each component in the proposed PSA-DSMS-TCN architecture, this study designed progressive ablation experiments. To ensure rigorous comparison, all baseline models uniformly adopted the Dual-Stream (DS) input structure and were independently optimized within the same hyperparameter space. The average performance metrics and statistical test results of each model under 5-fold cross-validation are presented in Table 2, and the 5-fold cross-validation accuracy comparison and confusion matrices are illustrated in Figure 8.

Table 2. Performance comparison: baseline vs. proposed model.

Model Architecture	Accuracy	Precision	Recall	F1-Score	<i>p</i> -Value	Cohen_D
DS-CNN	88.13% ± 4.11%	89.16%	88.79%	0.8808	0.064	1.33
DS-TCN	89.29% ± 3%	90.39%	89.6%	0.892	0.0384	1.23
DSMS-TCN	91.68% ± 2.18%	92.85%	91.69%	0.9157	0.1163	0.88
PSA-DSMS-TCN	94.05% ± 3.75%	94.47%	94.04%	0.9399	---	---

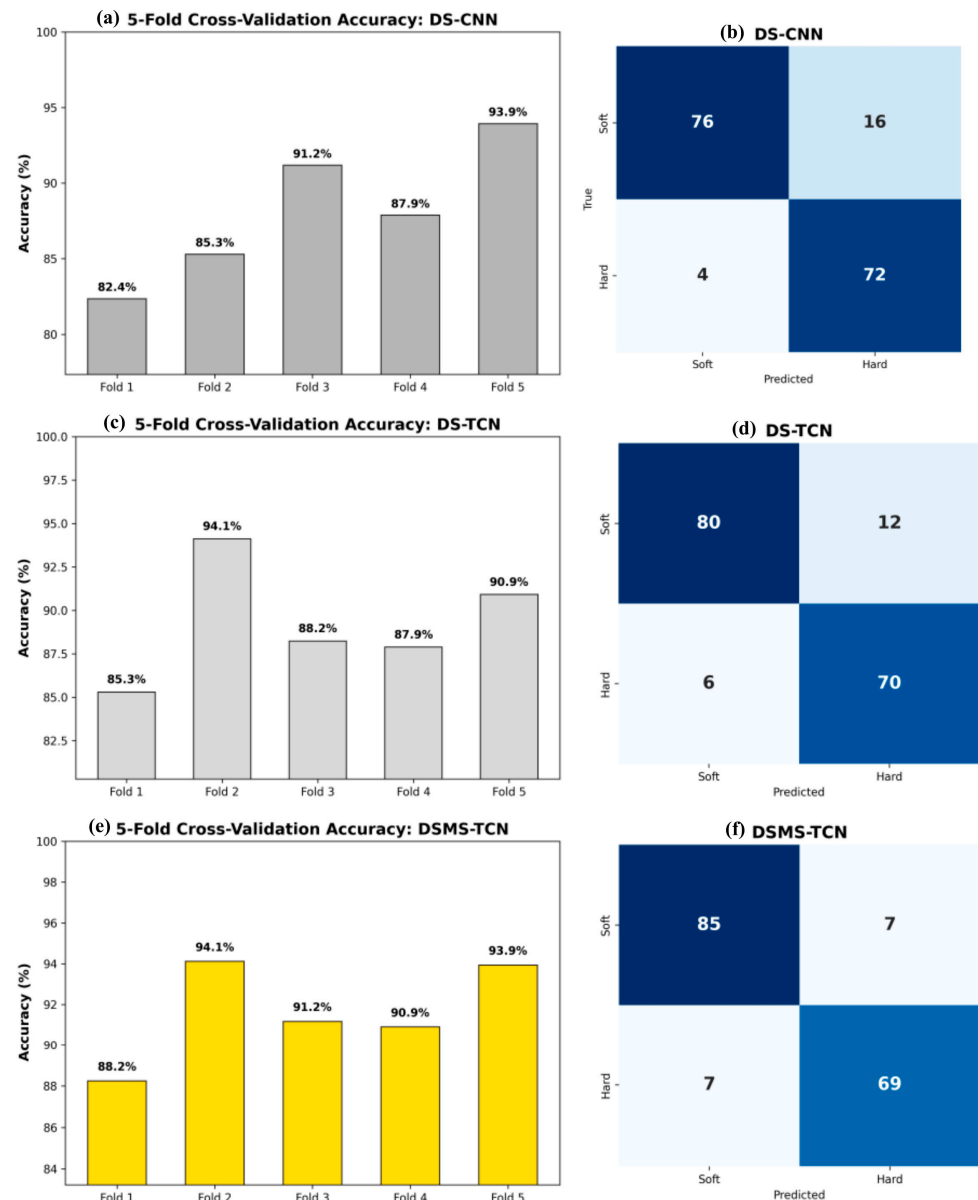


Figure 8. (a) 5-fold cross-validation accuracy comparison of the DS-CNN model. (b) Confusion matrix of the DS-CNN model. (c) 5-fold cross-validation accuracy comparison of the DS-TCN model. (d) Confusion matrix of the DS-TCN model. (e) 5-fold cross-validation accuracy comparison of the DSMS-TCN model. (f) Confusion matrix of the DSMS-TCN model.

To provide a rigorous benchmark for performance evaluation, the DS-TCN was implemented as an external baseline, strictly adhering to the architectural principles of the standard Temporal Convolutional Network (TCN) proposed by Bai et al. [12]. Specifically, it incorporates causal dilated convolutions to efficiently capture long-range temporal dependencies in sensor data. Given the limited existing literature specifically addressing clothing material identification in drum washing machines, we adopted this widely recognized state-of-the-art sequence modeling framework as a standard reference. Furthermore, considering the constrained computing resources of industrial microcontrollers, we adapted the original architecture by employing Batch Normalization and a streamlined dual-stream structure.

The experimental results show an obvious stepwise performance improvement trend. By comparing DS-CNN and DS-TCN, it can be seen that the temporal convolution-based architecture improves the accuracy by 1.16% and exhibits better robustness in the F1-score. After further introducing the multi-scale mechanism, the accuracy of DSMS-TCN

increases to 91.68%, demonstrating the superiority of multi-scale dilated convolution in capturing cross-scale physical features in washing signals, such as short-term impacts and long-term trends.

The proposed PSA-DSMS-TCN further improves the accuracy by 2.37 percentage points compared with DSMS-TCN. To rule out the randomness of the performance gain, paired-sample t-tests were conducted in this study. The results show that PSA-DSMS-TCN has a statistical advantage over the standard DS-TCN ($p = 0.0384 < 0.05$). Although the p -value between the PSA module and the multi-scale module is marginal (0.1163)—likely due to the constrained sample size—its Cohen's d effect size reaches 0.88, which is classified as a large effect according to statistical criteria. This suggests that the incremental contribution of the physics-aware mechanism to feature fusion is practically meaningful and offers robust engineering value.

From the confusion matrix of PSA-DSMS-TCN, it can be observed that the color is most concentrated on the main diagonal, indicating the lowest misclassification rate. In particular, the F1-score of the proposed model reaches 0.9399, significantly outperforming DS-CNN (0.8808). This implies that in practical washing scenarios, the model can more accurately balance the trade-off between soft load protection and hard load cleaning performance, effectively reducing the risk of mechanical damage caused by material misclassification, and thus possesses high industrial application value.

3. Conclusions

To address the urgent demand for accurate fabric material recognition in modern intelligent washing machines, this study investigates the mechatronic coupling mechanism during washing machine operation and proposes a deep learning solution incorporating physical prior knowledge. The core research and innovations are as follows: The physical feature mapping relationship in the loosening and weighing stages is established, clarifying the intrinsic correlation between vibration/current signals and clothing materials; the PSA-DSMS-TCN model is proposed, which alleviates heterogeneous signal interference through dual-stream decoupling and operating-condition gating architecture, thereby achieving efficient feature fusion. Experimental results show that the model reaches a 94.05% average accuracy and a 0.9399 F1-score during the pre-wash dry-sensing phase. These metrics confirm the model's robustness and effectiveness in capturing mechatronic features for material identification.

This method is implemented based on conventional configurations without additional sensors. The model has only 35.2 k parameters and an inference latency of 3.13 ms, enabling real-time judgment at zero time cost. Currently, this study focuses on binary classification. In future work, more material categories will be added based on the proposed framework, and the evolution laws of materials under wet-state conditions will be explored to further improve the generalization and recognition performance of the model under complex operating scenarios.

Author Contributions: Conceptualization, S.Z. and Y.G.; methodology and validation, Y.W. and B.Z.; writing—original draft, Y.G.; review and editing, S.Z. and J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Joint Research Program of the Yangtze River Delta Science and Technology Innovation Community (Grant No. 2024CSJGG1400), the National Natural Science Foundation of China (Grant No. 62471127), and the Suzhou Frontier Technology Research Project (Grant No. SYG202314).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: Authors Shurong Zhang and Bin Zhang were employed by the company Xiaomi Technology (Wuhan) Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Liu, Y.; Xie, S.; He, H.; Zhu, J. A Design for an Intelligent Washing Machine Control System. In *2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA)*; IEEE: Piscataway, NJ, USA, 2024; pp. 185–189.
2. Susto, G.A.; Vettore, L.; Zambonin, G.; Altinier, F.; Beninato, D.; Girotto, T.; Rampazzo, M.; Beghi, A. A Machine Learning-based Soft Sensor for Laundry Load Fabric Typology Estimation in Household Washer-Dryers. *IFAC-PapersOnLine* **2019**, *52*, 116–121. [\[CrossRef\]](#)
3. Pal, K. AI-Based One Touch Input Washing Machine. *Int. J. Res. Appl. Sci. Eng. Technol.* **2025**, *13*, 1587–1591. [\[CrossRef\]](#)
4. Ji, S.; Zhang, Z.; Stein, H.S.; Zhu, J. Flexible health prognosis of battery nonlinear aging using temporal transfer learning. *Appl. Energy* **2025**, *377*, 124766. [\[CrossRef\]](#)
5. Zhu, J.; Ji, S.; Ren, Z.; Wu, W.; Zhang, Z.; Ni, Z.; Liu, L.; Zhang, Z.; Song, A.; Lee, C. Triboelectric-induced ion mobility for artificial intelligence-enhanced mid-infrared gas spectroscopy. *Nat. Commun.* **2023**, *14*, 2524. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Bono, F.M.; Radicioni, L.; Cinquemani, S.; Conese, C.; Tarabini, M. Development of soft sensors based on neural networks for detection of anomaly working condition in automated machinery. In *Proceedings of the NDE 4.0, Predictive Maintenance, and Communication and Energy Systems in a Globally Networked World*, Long Beach, CA, USA, 6–10 March 2022; Volume 12049, pp. 56–70.
7. Kang, C.; Bae, G.; Kim, D.; Lee, K.; Son, D.; Lee, C.; Lee, J.; Lee, J.; Yun, J.W. Data-Driven Reinforcement Learning for Optimal Motor Control in Washing Machines. In *2024 IEEE Conference on Artificial Intelligence (CAI)*; IEEE: Piscataway, NJ, USA, 2024; pp. 418–424.
8. Xu, Z.; Zhao, K.; Wang, J.; Bashir, M. Physics-informed probabilistic deep network with interpretable mechanism for trustworthy mechanical fault diagnosis. *Adv. Eng. Inform.* **2024**, *62*, 102806. [\[CrossRef\]](#)
9. Ji, S.; Zhu, J.; Lyu, Z.; You, H.; Zhou, Y.; Gu, L.; Qu, J.; Xia, Z.; Zhang, Z.; Dai, H. Deep learning enhanced lithium-ion battery nonlinear fading prognosis. *J. Energy Chem.* **2023**, *78*, 565–573. [\[CrossRef\]](#)
10. Praticò, D.; Laganà, F. Infrared Thermographic Signal Analysis of Bioactive Edible Oils Using CNNs for Quality Assessment. *Signals* **2025**, *6*, 38. [\[CrossRef\]](#)
11. Hao, Z.; Liu, S.; Zhang, Y.; Ying, C.; Feng, Y.; Su, H.; Zhu, J. Physics-Informed Machine Learning: A Survey on Problems, Methods and Applications. *arXiv* **2023**, arXiv:2211.08064. [\[CrossRef\]](#)
12. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv* **2018**, arXiv:1803.01271. [\[CrossRef\]](#)
13. Zhao, W.; Gao, Y.; Ji, T.; Wan, X.; Ye, F.; Bai, G. Deep Temporal Convolutional Networks for Short-Term Traffic Flow Forecasting. *IEEE Access* **2019**, *7*, 114496–114507. [\[CrossRef\]](#)
14. Martinez, B.; Ma, P.; Petridis, S.; Pantic, M. Lipreading Using Temporal Convolutional Networks. In *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; IEEE: Piscataway, NJ, USA, 2020; pp. 6319–6323.
15. He, Y.; Zhao, J. Temporal Convolutional Networks for Anomaly Detection in Time Series. *J. Phys. Conf. Ser.* **2019**, *1213*, 042050. [\[CrossRef\]](#)
16. Lara-Benítez, P.; Carranza-García, M.; Luna-Romera, J.M.; Riquelme, J.C. Temporal Convolutional Networks Applied to Energy-Related Time Series Forecasting. *Appl. Sci.* **2020**, *10*, 2322. [\[CrossRef\]](#)
17. van den Oord, A.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K. WaveNet: A Generative Model for Raw Audio. *arXiv* **2016**, arXiv:1609.03499. [\[CrossRef\]](#)
18. Lin, G.; Wu, Q.; Qiu, L.; Huang, X. Image super-resolution using a dilated convolutional neural network. *Neurocomputing* **2018**, *275*, 1219–1230. [\[CrossRef\]](#)
19. Ahmed, K.R. Smart Pothole Detection Using Deep Learning Based on Dilated Convolution. *Sensors* **2021**, *21*, 8406. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Diversi, R.; Lenzi, A.; Speciale, N.; Barbieri, M. An Autoregressive-Based Motor Current Signature Analysis Approach for Fault Diagnosis of Electric Motor-Driven Mechanisms. *Sensors* **2025**, *25*, 1130. [\[CrossRef\]](#)
21. Yang, J.; Gao, T.; Zhang, H.; Li, Y. A multi-sensor fault diagnosis method for rotating machinery based on improved fuzzy support fusion and self-normalized spatio-temporal network. *Meas. Sci. Technol.* **2023**, *34*, 125112. [\[CrossRef\]](#)

22. Kibrete, F.; Engida Woldemichael, D.; Shimels Gebremedhen, H. Multi-Sensor data fusion in intelligent fault diagnosis of rotating machines: A comprehensive review. *Measurement* **2024**, *232*, 114658. [[CrossRef](#)]
23. Wang, Y.; Wang, G.; Chen, C.; Pan, Z. Multi-scale dilated convolution of convolutional neural network for image denoising. *Multimed. Tools Appl.* **2019**, *78*, 19945–19960. [[CrossRef](#)]
24. Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. *arXiv* **2016**, arXiv:1511.07122. [[CrossRef](#)]
25. Ma, R.; Han, T.; Lei, W. Cross-domain meta learning fault diagnosis based on multi-scale dilated convolution and adaptive relation module. *Knowl.-Based Syst.* **2023**, *261*, 110175. [[CrossRef](#)]
26. Xue, L.; Li, X.; Zhang, N.L. Not All Attention Is Needed: Gated Attention Network for Sequence Data. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 6550–6557. [[CrossRef](#)]
27. Sangwan, S.; Chauhan, D.S.; Akhtar, M.S.; Ekbal, A.; Bhattacharyya, P. Multi-task Gated Contextual Cross-Modal Attention Framework for Sentiment and Emotion Analysis. In *Neural Information Processing*; Gedeon, T., Wong, K.W., Lee, M., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 662–669.
28. Chen, D.; Huang, T.; Song, Z.; Deng, S.; Jia, T. AGG-Net: Attention Guided Gated-Convolutional Network for Depth Image Completion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–8 October 2023*; pp. 8853–8862.

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