

Dynamic Feature Extraction and Interactive Visualization: A Human-Collaborative Intelligent Analysis System for Complex Textile Manufacturing Datasets

Ying Pan

Department of Computer Science and Technology, School of Artificial Intelligence, Harbin University, Harbin 150086, Heilongjiang, China
Corresponding author: Ying Pan (e-mail: m13936101481@163.com)

ABSTRACT As the era of big data continues to evolve, the proliferation of massive, multi-dimensional datasets—particularly those generated by high-speed fiber manufacturing and intelligent textile monitoring systems—poses significant challenges to traditional data analysis methodologies. In modern intelligent sensing environments, electromagnetic signal acquisition and wireless monitoring technologies further increase the complexity and dynamic characteristics of industrial data streams, requiring more adaptive analytical frameworks. This paper introduces a novel approach that integrates machine learning-based feature extraction with interactive dynamic visualization to establish a tightly coupled human-collaborative intelligent analysis system. The proposed framework transforms conventional static feature extraction into a dynamic iterative process encompassing feature perception, model optimization, insight verification, dynamic feature engineering, importance-shift-based visualization mapping, and real-time human-machine interaction feedback. Application examples involving complex textile manufacturing datasets demonstrate that the framework not only improves the efficiency and interpretability of feature extraction for fiber quality parameters but also enhances analysts' understanding of model behavior and decision-making logic. The results indicate broad application prospects for intelligent industrial monitoring and adaptive data analysis while providing useful references for electromagnetic sensing-assisted information processing and real-time decision support in complex manufacturing environments.

INDEX TERMS feature extraction, dynamic visualization, human-computer collaboration, electromagnetic sensing, intelligent data analytics

I. INTRODUCTION

IN the contemporary information age, decision-making increasingly relies on the extraction of actionable insights from vast and complex datasets. The high dimensionality, multimodal nature, and rapid evolution of data streams generated by automated spinning and weaving sensors render conventional statistical techniques inadequate for real-time analysis and visualization of multifaceted textile manufacturing variables. These unstructured datasets, which fluctuate in response to yarn quality and machine performance, necessitate advanced computational frameworks capable of capturing intricate correlations inherent in modern textile production environments. Deep learning and machine learning techniques have made significant strides in automating feature extraction and learning discriminative, high-level representations from raw data, leading to breakthroughs across numerous applications. However, current technolog-

ical practices often treat feature extraction as a singular, isolated preprocessing step, resulting in reduced flexibility, adaptability, and interactivity during the deployment of machine learning models, with extracted features remaining static. To address these limitations, this paper proposes a paradigm that fully integrates automated feature extraction and profiling via machine learning with dynamically interactive feature analysis and display through visualization. The objective is to dissolve the divide between feature extraction and result presentation, establishing a real-time bidirectional interactive feedback loop. Within this loop, visualization serves not only as a conduit for presenting final results but also as a mechanism for adaptive model optimization and enhanced interpretability. This paper details the technical architecture and procedures required to dynamically extract features from complex data, represent feature and model states in real time, and enable iterative

feature learning through human-computer interaction. The significance of this research lies in the advancement of proactive data visualization techniques that offer intelligent analytical support for managing the high-dimensional sensor data characteristic of fiber production and weaving processes, transforming complex textile manufacturing datasets into powerful, accessible analytical systems. By integrating real-time informational feedback, the framework facilitates a more intuitive progression from raw industrial variables to actionable insights, thus improving the precision and usability of decision-making systems in modern textile engineering.

II. RELATED WORK

Within the data-driven research paradigm, efficient and automated extraction of physically meaningful and discriminative features from increasingly complex, high-dimensional data has emerged as a critical technological bottleneck for advancing intelligence across various domains. Li et al. [1] employed diverse feature extraction and selection methods to construct and identify features most predictive of TBM performance from multi-source time series data. Abidi et al. [2] reviewed the evolution of facial feature extraction technologies, spanning from traditional manual techniques to state-of-the-art deep learning approaches. Mayorca-Torres et al. [3] summarized spatiotemporal, graphical, and image feature extraction methods derived from ECGI data. Kunang et al. [4] utilized a deep autoencoder based on unsupervised learning to extract features from raw or preprocessed IoT network traffic data. Zhao et al. [5] developed and released the environmental microbial image dataset EMDS-6 to systematically evaluate and enhance the performance of computer vision tasks including image denoising, semantic segmentation, feature extraction, image classification, and object detection in environmental microbial analysis.

Telmem et al. [6] transformed speech signals into Mel spectrograms as input features. For periodic physiological signals such as ECG and PPG, Bondugula et al. [7] applied the Miniature Random Convolutional Kernel Transform (MINIROCKET) method to rapidly extract numerous features. Fiscione et al. [8] curated a multiparameter MRI dataset specifically for radiomics analysis of magnetic susceptibility imaging. Osuna-González et al. [9] leveraged bioinformatics tools and processes to standardize genomic data and extract key genomic features. Aleesa et al. [10] proposed an efficient feature extraction method for the GFER task, generating feature vectors highly sensitive to grammatical facial movements with minimal computational overhead, subsequently facilitating combination recognition. However, most existing studies treat feature extraction as a static, offline preprocessing step, lacking capabilities for real-time adaptation to dynamic data streams, feature importance drift, and human-computer interaction feedback. The framework proposed herein seeks to transcend the traditional linear process from feature extraction to model input. Its innovation lies in the deep integration of machine learning's

automatic feature extraction capabilities with the dynamic display and reasoning functions of interactive visualization, forming a closed-loop, adaptive system.

III. METHODS

A. INTELLIGENT FEATURE EXTRACTION AND DYNAMIC FEATURE ENGINEERING

The foundation of the system is feature extraction, employing a hierarchical, hybrid strategy to assimilate data from diverse sources. For structured numerical data, the LightGBM feature importance evaluation method is utilized alongside multi-dimensional feature selection metrics such as mutual information. For time-series data, traditional statistical features (mean, variance, trend) are complemented by efficient convolutional algorithms like MINIROCKET to rapidly generate numerous informative time-series features. In the context of image data, feature descriptors are derived from the activation of intermediate layers in ResNet models. The centerpiece of this module is dynamic feature engineering, which maintains a dynamic feature pool rather than a static feature set. Feature significance is determined in real time by an online learner, which continuously evaluates the contribution of each feature to the model's predictive target, calculating importance scores that evolve over time. To distinguish between feature importance shift from the model perspective and concept drift in the underlying data, the system concurrently monitors changes in prediction errors and feature importance distributions. Only when shifts in feature importance coincide with potential degradation in model performance are they attributed to concept drift, prompting more substantial adjustments to the feature pool. As new data batches arrive, the system adds or removes features from the feature pool based on updated importance scores: persistently low-importance features are assigned lower weights or eliminated, while emerging patterns in new data prompt the creation and evaluation of new features.

To preserve model stability and prevent excessive prediction variance, feature pool updates follow a smooth transition strategy: the introduction of new features and removal of obsolete ones occur gradually over several time windows through weighted fusion. Additionally, the system establishes a stable zone for core features, which are downgraded only after sustained declines in importance are verified as redundancy. Deleted features are not permanently discarded; instead, a candidate feature archive is maintained, where archived features are periodically re-evaluated using new data samples. If a previously removed feature regains significance, it is reactivated and reintroduced into the dynamic feature pool. This mechanism ensures the feature set adapts to changes in data distribution [11,12]. Table 1 illustrates variations in feature importance scores and corresponding system operations across four successive time windows:

TABLE 1. Importance Assessment and Operational Simulation of Dynamic Feature Pool

Feature	Time Window T1	Time Window T2	Time Window T3	Time Window T4	Description of Characteristic State Evolution
F1	0.85	0.92	0.88	0.45	Key feature importance decreases markedly at T4, possibly indicating operational changes or feature failure.
F2	0.72	0.68	0.15	0.10	Importance continuously declines; system automatically downgrades after T3, potentially removing it from the main feature set.
F3	–	0.30	0.65	0.82	Introduced as a new feature in T2; importance rises steadily, becoming a core feature by T4.
F4	0.50	0.55	0.52	0.51	Importance remains stable, reflecting persistent correlation.
F5	0.08	0.12	–	–	Low initial importance; system temporarily removes after T2.

Table 1 visualizes the data processing within the dynamic feature engineering module. The online evaluator updates feature importance scores at the conclusion of each time window. For example, feature F1 maintains a high score from T1 through T3, indicating strong discriminative power, but its score drops significantly in T4, prompting the system to reduce its model training and visualization weight to accommodate potential changes in data distribution. Feature F3, introduced in T2, demonstrates increasing significance, ultimately becoming a core feature by T4. In contrast, feature F5 is proactively removed due to its consistently low importance, enhancing computational efficiency and model focus. This adaptive process illustrates the system's capacity to address concept drift and self-optimize the feature set in response to evolving data distributions via a dynamic rather than fixed feature pool.

B. COMPUTATIONAL EFFICIENCY AND TIME COMPLEXITY ANALYSIS

To ensure real-time applicability in environments such as high-speed textile production, the framework specifically optimizes key algorithms. For MINIROCKET feature extraction, its inherent parallelization and linear time complexity ($O(n)$) enable rapid processing of incoming time-series data blocks. To mitigate the high computational cost of t-SNE, an incremental, approximate t-SNE algorithm is employed, performing dimensionality reduction solely on representative sample subsets when significant changes in feature importance distributions occur or upon user request. This reduces computational complexity from $O(n^2)$ to approximately $O(k_n)$, where k denotes the sample subset size. SHAP value calculation is similarly approximated, focusing detailed analysis on key features and sample points rather than exhaustive computation. Under experimental hardware conditions (Intel Xeon 8-core processor, 32 GB RAM), for each batch of sensor data streams comprising approximately 100,000 data points across 50 dimensions, end-to-end system latency—from feature extraction and dynamic updates to visual rendering—averaged less than 2 seconds, satisfying the interactive requirements of most industrial real-time monitoring scenarios.

C. INTERACTIVE VISUAL MAPPING AND DYNAMIC DISPLAY DESIGN

The visualization mapping engine presents information from the dynamic feature pool and the internal state of the machine learning model to users in an intuitive and interactive manner. For high-dimensional feature spaces, an improved t-SNE algorithm is utilized, with input consisting of weighted feature vectors where weights are dynamically calculated importance scores [13]. This approach ensures that dimensionality reduction directly reflects the current data distribution as defined by the most significant features. When feature importance shifts, the visualization layout transitions smoothly through animation, visually depicting the evolution of data concepts.

The system integrates multiple views. The primary view displays data point distributions in dynamic feature space. Another critical perspective is the Feature River Plot, which depicts the temporal evolution and magnitude of feature importance, revealing the timing and extent of concept drift. For time-series data, overlay plots co-locate time-series curves with key feature points (e.g., extremes and change points) extracted from the data. Model interpretability is further enhanced through SHapley Additive exPlanations (SHAP) value waterfall charts, enabling users to understand the rationale behind model decisions for specific data points.

D. REAL-TIME FEEDBACK LOOP AND MODEL OPTIMIZATION MECHANISM

The interactive feedback loop serves as the interface between users and the system, facilitating intelligent human-machine collaboration. Users interact via a visual interface, performing tasks that generate real-time feedback indicators for feature extraction and prediction models. For instance, users may annotate misclassified data points or those of particular interest on a scatter plot; these labels are immediately incorporated as supervisory signals by the online learning algorithm to optimize feature extractor parameters or adjust classifier decision boundaries.

Feature-level intervention is also supported. Users can increase the weight of features undervalued by the system or suppress those identified as noise, leveraging domain expertise. This intervention is operationalized by modifying the online learner's loss function: for features designated for in-

creased weight, a regularization term is added, inversely proportional to the discrepancy between the model-assigned and user-specified weights, thereby guiding subsequent model updates toward user expectations. To prevent excessive influence from individual interventions, a confidence attenuation coefficient is implemented, and statistical information from multiple interaction rounds or historically successful interventions is integrated to maintain learning robustness. These interventions directly affect subsequent visualization mapping and model training inputs. The system records user interaction history, forming implicit feedback that optimizes default visualization configurations and progressively adapts to user analysis patterns. Through this real-time feedback loop, domain knowledge is effectively embedded into the automated machine learning process, establishing a virtuous cycle of hypothesis, verification, and adjustment, which continually enhances analytical accuracy and insight depth [14].

IV. RESULTS AND DISCUSSION

A. APPLICATION AND ANALYSIS ON

HIGH-DIMENSIONAL INDUSTRIAL SENSOR DATASETS

The experimental dataset comprises multi-source sensor data collected from industrial machinery via Internet of Things installations, containing time-series records from 50 sensors

over three months, along with equipment fault labels. This high-dimensional data exhibits intricate correlations among sensors, and equipment performance degradation induces gradual shifts in data distribution.

Employing the proposed framework, the dynamic feature engineering module extracted over 200 time-series and statistical features from the raw sensor data. Real-time monitoring of feature importance revealed pronounced changes in the set of key features during normal equipment operation and the onset of failures. For example, the center of gravity attribute of vibration sensor A's spectrum exhibited low importance during normal operation but increased sharply upon initial failure conditions, becoming a critical early warning indicator. Table 2 compares fault prediction performance on the test set using static feature extraction (based on the preceding 30 days, with a fixed feature set) and the dynamic feature extraction method under the proposed framework. Two static baselines were established for fair comparison:

- (a) Static-Once (model trained once using the first 30 days of data and fixed thereafter);
- (b) Static-Retrain (model retrained every 30 days with a fixed feature set and new data).

Evaluation metrics include accuracy, precision, and recall.

TABLE 2. Comparison of Industrial Sensor Data Fault Prediction Performance

Method	Accuracy (%)	Precision (%)	Recall (%)
Static-Once (Model trained once and fixed)	68.5	65.2	62.8
Static-Retrain (Model retrained every 30 days with fixed feature set)	79.3	76.1	74.5
Dynamic Feature Extraction Method (Proposed Framework)	92.7	89.4	88.6

The dynamic feature extraction method significantly outperformed static methods across all evaluation metrics. Its ability to promptly capture data distribution drift and adjust feature weights maintains high sensitivity to emerging failure modes. Even the Static-Retrain model, which is periodically updated, cannot flexibly introduce new features or eliminate obsolete ones, resulting in delayed performance. The Feature River Map visualization depicts the temporal evolution of sensor feature importance, providing maintenance personnel with a clear representation of equipment health degradation and enhancing interpretability and decision support.

B. DISCOVERY OF DYNAMIC PATTERNS IN TIME-SERIES FINANCIAL TRANSACTION DATA

The second experiment targeted high-frequency financial trading data to identify anomalous trading patterns. The dataset consisted of one month's transaction records for a single stock, including time, price, and volume fields.

The framework's time-series feature extraction module derived features such as the slope of the volume-weighted average price (VWAP), the frequency of large orders, and

price volatility clustering from the trading sequence. Dynamic visualization mapped data points (time windows) onto a two-dimensional plane, enabling analysts to identify clusters corresponding to periods of sudden market liquidity depletion. SHAP analysis showed that the key features driving cluster separation were sharp declines in order book depth and sudden widening of the bid-ask spread.

Using the feedback loop, analysts labeled these clusters as liquidity risk events. The system subsequently highlighted and warned of trading windows exhibiting similar patterns in real-time data streams. Table 3 compares system-generated alerts to actual risk events manually identified by senior analysts in the following week's data.

TABLE 3. Early Warning Effect of Abnormal Financial Transaction Patterns

Project	Quantity
Total number of system warnings	15
Total number of real risk events	12
Number of events successfully alerted by the system	11
False alarm count	4
Number of missed reports	1
Early warning accuracy	73.3%
Early warning recall rate	91.7%

The system achieved a high recall rate, successfully detecting most risk events that are challenging for humans to identify in real time, despite some false positives. This demonstrates the framework’s capability to discover and track emerging patterns in dynamic, rapidly evolving time-series data. Dynamic visualization empowers analysts to comprehend not only the existence of anomalous patterns but also the underlying causes.

C. EVALUATION OF INTERPRETABILITY ENHANCEMENT ON IMAGE DATASETS

The third experiment utilized a publicly available medical image dataset for dermoscopy image classification, aiming to distinguish benign from malignant skin lesions. Feature extraction was performed using a pre-trained convolutional neural network (CNN), with visualization of deep feature spatial distributions provided dynamically.

The experiment focused on evaluating improvements in

model interpretability. Fifteen domain experts were asked to assess two interpretation methods:

- (1) the traditional approach, presenting classification results and confidence scores;
- (2) the proposed framework, presenting a feature space distribution map and interpretability measures (SHAP plots) for specific images.

Experts rated comprehensibility (1–5 points; higher scores indicate greater clarity) and trustworthiness (1–5 points; higher scores indicate greater confidence). To minimize visual appeal bias, both methods employed interfaces with unified design and comparable information density, and evaluation order was randomized. Additionally, an objective metric was introduced: experts predicted model classifications for challenging edge-case images based on the provided explanations, measuring the accuracy of decision logic conveyed.

TABLE 4. Expert Evaluation Results of Model Interpretability

Evaluation Project	Results Only	Visualization Framework (Proposed)	Increase
Average Comprehension Score	2.2	4.4	100.0%
Average Trust Score	2.6	4.2	61.5%
Decision Logic Prediction Accuracy	38%	82%	115.8%

The results demonstrate that the dynamic visualization framework substantially enhances experts’ understanding and trust in model decisions. The marked increase in decision logic prediction accuracy confirms that the framework effectively communicates the model’s internal reasoning, independent of interface attractiveness. Experts reported that visualization enabled them to observe how the model bases its judgments on features in specific image regions, which aligns with their clinical diagnostic processes and fosters trust in AI applications. Such white-box interpretability is essential for the adoption of AI in high-risk domains such as healthcare.

Across all three experiments, the core strengths of the proposed framework lie in its dynamic and interactive nature. It functions not only as an advanced feature extractor but also as a collaborative system supporting exploratory data analysis. By integrating human expertise into the

machine learning loop, the framework adapts to changing environments, uncovers novel patterns, and enhances trust in complex models, establishing a new paradigm for managing complex datasets.

V. CONCLUSION

This paper addresses the challenges of non-interactive, opaque analysis of complex textile manufacturing data by presenting a technical architecture that merges machine learning-based feature extraction with interactive dynamic visualization, elucidating the intricate relationships between fiber quality parameters and production variables. This integration transforms automated monitoring outputs into interpretable visual insights, enabling more precise operational control and fault diagnosis in high-speed textile machinery environments. The architecture converts the traditional linear data processing pipeline into a human-computer interactive,

bidirectional analytical system comprising three modules: intelligent feature extraction, dynamic visualization mapping, and a real-time feedback loop. Future research directions include:

- (1) enhancing algorithmic efficiency in feature engineering to support ultra-large-scale data processing;
- (2) developing richer, more natural interaction modalities (e.g., natural language queries and virtual reality environments) to lower system entry barriers;
- (3) investigating multimodal data fusion scenarios to evaluate the system's effectiveness in correlating diverse data types and identifying cross-modal patterns.

AUTHOR CONTRIBUTIONS

Ying Pan designed, collected and analyzed the data, and drafted the manuscript. Ying Pan conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Ying Pan participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] J. B. Li, Z. Y. Chen, X. Li, et al., "Feedback on a shared big dataset for intelligent TBM Part I: Feature extraction and machine learning methods," *Underground Space*, vol. 11, pp. 1-25, 2023, doi: 10.1016/j.undsp.2023.01.001.
- [2] S. M. H. Abidi, S. A. Hassan, S. M. Raza, et al., "Advances in Face Recognition: A Comprehensive Review of Feature Extraction and Dataset Evaluation," *Electronics*, vol. 15, no. 2, pp. 338, 2026, doi: 10.3390/electronics15020338.
- [3] D. Mayorca-Torres, A. J. León-Salas, and D. H. Peluffo-Ordoñez, "Systematic review of computational techniques, dataset utilization, and feature extraction in electrocardiographic imaging," *Medical & Biological Engineering & Computing*, vol. 63, no. 5, pp. 1289-1317, 2025, doi: 10.1007/s11517-024-03264-z.
- [4] Y. N. Kunang, S. Nurmaini, D. Stiawan, et al., "An end-to-end intrusion detection system with IoT dataset using deep learning with unsupervised feature extraction," *International Journal of Information Security*, vol. 23, no. 3, pp. 1619-1648, 2024, doi: 10.1007/s10207-023-00807-7.
- [5] P. Zhao, C. Li, M. M. Rahaman, et al., "Emds-6: Environmental microorganism image dataset sixth version for image denoising, segmentation, feature extraction, classification, and detection method evaluation," *Frontiers in Microbiology*, vol. 13, Art. no. 829027, 2022, doi: 10.3389/fmicb.2022.829027.
- [6] M. Telmem, N. Laaidi, Y. Ghanou, et al., "Comparative study of CNN, LSTM and hybrid CNN-LSTM model in amazigh speech recognition using spectrogram feature extraction and different gender and age dataset," *International Journal of Speech Technology*, vol. 27, no. 4, pp. 1121-1133, 2024, doi: 10.1007/s10772-024-10154-0.
- [7] R. K. Bondugula, S. K. Udgata, and K. B. Sivangi, "A novel deep learning architecture and MINIROCKET feature extraction method for human activity recognition using ECG, PPG and inertial sensor dataset," *Applied Intelligence*, vol. 53, no. 11, pp. 14400-14425, 2023, doi: 10.1007/s10489-022-04250-4.
- [8] C. Fiscione, G. Sighinolfi, D. N. Manners, et al., "Multiparametric MRI dataset for susceptibility-based radiomic feature extraction and analysis," *Scientific Data*, vol. 11, no. 1, pp. 575, 2024, doi: 10.1038/s41597-024-03418-6.
- [9] R. Osuna-González, G. D. I. Luna, R. M. Valdovinos-Rosas, et al., "Structured RDF dataset for genomic feature extraction and detection of biotechnological microorganisms of the Burkholderia genus," *Data in Brief*, Art. no. 112137, 2025, doi: 10.1016/j.dib.2025.112137.
- [10] R. S. Aleesa, H. M. Mohammadi, A. Monadjemi, et al., "Dataset classification: An efficient feature extraction approach for grammatical facial expression recognition," *Computers and Electrical Engineering*, vol. 110, Art. no. 108891, 2023, doi: 10.1016/j.compeleceng.2023.108891.
- [11] S. Dass and P. Jayagopal, "A novel hybrid feature extraction methods for eeg based emotions classification using ensemble approaches in hied dataset," *Recent Advances in Computer Science and Communications*, vol. 18, no. 5, Art. no. E280624231442, 2025, doi: 10.2174/0126662558306769240624053833.
- [12] S. Dua, S. Kumar, R. Garg, et al., "Residual network-based feature extraction for automatic crop disease detection system using drone image dataset," *International Journal of Intelligent Unmanned Systems*, vol. 13, no. 1, pp. 54-77, 2025, doi: 10.1108/IJIUS-08-2024-0248.
- [13] X. N. Kee and I. R. A. Hamid, "Feature Extraction Tool for Phishing Dataset," *Applied Information Technology And Computer Science*, vol. 4, no. 2, pp. 248-265, 2023, doi: 10.30880/aitcs.2023.04.02.015.
- [14] D. Javed, U. Arshad, H. Irfan, et al., "Evaluation of Boosting Algorithms for Skin Cancer Classification Using the PAD-UFES-20 Dataset and Custom CNN Feature Extraction," *Engineering Proceedings*, vol. 87, no. 1, pp. 115, 2025, doi: 10.3390/engproc2025087115.