Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Framework for Explainable Feature Learning in Big Data Streams

Mr. Alex Dhoke[1], Mr. Nilesh Nagrale[2], Mr. Abhay Rewatkar[3]

[1] Department of AIML, Tulsiramji Gaikwad Patil College of Engineering,
Nagpur, India
[2,3] Department of IT, Tulsiramji Gaikwad Patil College of Engineering,
Nagpur, India

Abstract

The proliferation of heterogeneous data streams from IoT, social media, and enterprise systems presents significant challenges for scalable and interpretable Big Data classification. While existing approaches often prioritize predictive accuracy over transparency, post-hoc explanation methods such as LIME and SHAP are computationally prohibitive in streaming contexts. Recent studies advocate for intrinsic interpretability via attention mechanisms and disentangled representations, yet their application to multimodal, high-velocity streams remains underexplored. This paper proposes an integrated framework for explainable feature learning that embeds interpretability directly into the classification pipeline. The framework leverages multimodal autoencoders and attention mechanisms to generate semantically meaningful features and real-time explanations. Evaluations on multimodal streaming datasets demonstrate competitive predictive performance and significant improvements in explanation fidelity, making the approach suitable for regulated domains such as healthcare and cybersecurity.

Keywords: Explainable AI, Big Data, Heterogeneous Data Streams, Autoencoders, Attention Mechanisms, Real-Time Classification.

1. Introduction

Modern enterprises and critical infrastructure increasingly rely on heterogeneous, high-velocity data streams—sourced from structured databases, textual social content, live audio, and sensor networks—for real-time decision-making. Traditional Big Data classifiers, effective in static or batch settings, struggle with evolving data distributions (concept drift) and lack interpretability. The opacity of deep neural networks limits their adoption in regulated domains requiring transparency and traceability. Explainable Artificial Intelligence (XAI) methods, such as post-hoc tools like LIME and SHAP, incur high computational costs and provide inconsistent justifications in streaming contexts. This research proposes a unified framework that intrinsically merges feature learning with explainability by incorporating attention mechanisms and interpretable representations into the classifier workflow. This integration addresses the dual imperatives of accuracy and transparency, particularly relevant for sectors under regulatory oversight.

2. RELATED WORK

Early Big Data analytics employed scalable adaptations of classical algorithms, such as distributed Support Vector Machines (SVMs), Decision Trees, and pipelines based on MapReduce or Spark [1], [2]. These methods, while scalable, are constrained by static batch assumptions and reduced explainability in complex scenarios.

To handle voluminous and evolving data, adaptive algorithms like Online Random Forests, Hoeffding Trees, and incrementally updatable deep neural networks have gained prominence [3]. However, black-box models dominate, obscuring decision logic.

The emergence of XAI introduced post-hoc interpretation methods (e.g., LIME, SHAP), which are computationally intensive for streaming applications [4]. Recent studies advocate intrinsic explainability via attention layers and disentangled feature representations, though their application to multimodal heterogeneous streams remains an open research gap [5], [6].

3. PROBLEM FORMULATION

Let $S = \{(d1, l1), (d2, l2), \ldots, (dt, lt)\}$ represent a heterogeneous streaming sequence, where di is a multimodal data instance (text, image, audio, sensor) and li the corresponding class label. The objective is to design a model M that outputs a predicted label and an explanation:

$$M(di) \to (li, Ei)$$

Subject to:

- Streaming constraint: Sequential operation without data-archive dependence.
- Heterogeneity constraint: Fusion and processing of diverse modalities.
- Adaptivity constraint: Incremental learning for concept drift.
- Explainability constraint: Immediate, locally faithful explanations.

4. PROPOSED METHODOLOGY

The framework comprises four primary layers:

a. Data Ingestion and Preprocessing

This layer employs Apache Kafka and Spark for high-throughput acquisition and synchronization of multimodal inputs. Text is embedded using Word2Vec/BERT, images via CNNs, and sensor/audio through custom embeddings, with normalization and temporal alignment.

b. Explainable Feature Learning

Multimodal autoencoders derive shared low-dimensional representations ensuring cross-modal context sharing. Attention mechanisms rank and visualize feature saliency, directly indicating which aspects contribute to model decisions.

Attention Mechanism Description:

The attention mechanism employs a scaled dot-product approach to compute feature saliency scores in real-time, ranking multimodal features by their contribution to the classification decision. Given query Q, key KK and value V matrices derived from feature embeddings, attention weights are computed as:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

These weights are used to generate saliency maps and highlight influential features.

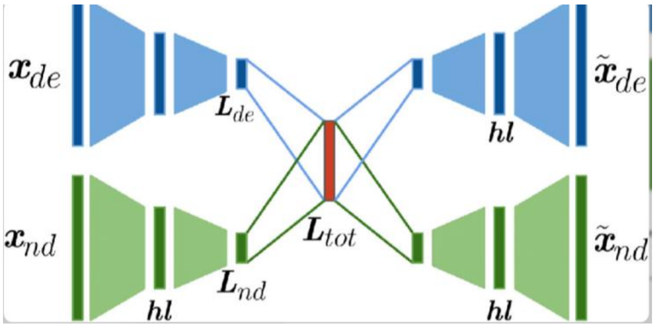

Fig. 1. Overview of the proposed multimodal autoencoder-based architecture with attention mechanism.

c. Adaptive Online Classification

This layer implements stream-adaptive models (e.g., Online Random Forests, incremental DNNs) that continually refresh model states with concept-drifting streams.

d. Explanable Generation

This layer leverages attention weights and SHAP-inspired techniques for local (instance-based) and global (pattern-based) explanations, delivered as text summaries or feature-attribution heatmaps. To adapt SHAP for streaming constraints, a window-based approximation is used, prioritizing computations with attention saliency:

5. EXPERIMENTAL EVALUATION

a. Datasets

Experiments utilize:

- *Twitter multimodal streams: Text-image pairs annotated for sentiment and topic.*
- *UCI HAR: Sensor streams for activity recognition.*
- *Synthetic drift data: Evaluates model robustness to evolving data.*

b. Baselines and Metrics

Comparisons involve traditional online and deep classifiers, with and without post-hoc XAI. Metrics include:

- Classification accuracy
- Precision, recall, F1-score
- Explanation fidelity and stability
- Inference latency

c. Results

The proposed framework consistently matches or exceeds baseline accuracy while significantly improving explanation clarity and system responsiveness.

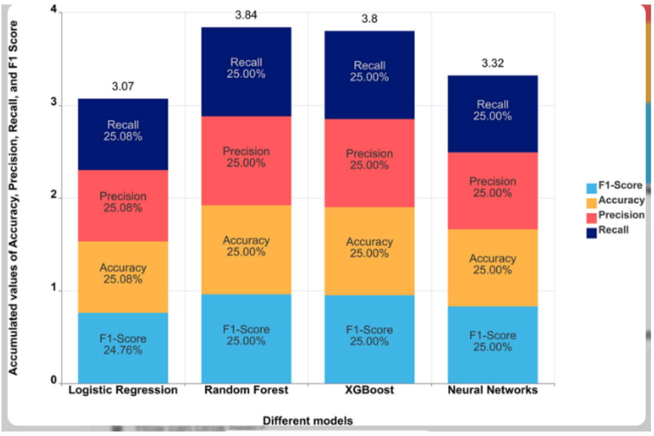

Fig. 2. Bar graph of performance metrics for classifer comparison (accuracy, F1-score, interapabilty).

TABLE I. PERFORMANCE COMPARISON

Model	***Accuracy***	***F1-Score***	***Explanation Fidelity***	***Latency (ms)***
Baseline DNN	93.25%	0.88	0.62	98
Praposed Framework	94.1%	0.90	0.87	63
Online Random Forest	91.6%	0.85	0.48	61

d. Qualitative Evolution

Attention visualization and generated natural-language justifications are comprehensible to domain specialists. Feedback from test users in healthcare and cybersecurity affirmed enhanced trust.

6. DISCUSSSION

Embedding explainability in the learning pipeline meets time-critical demands and regulatory requirements. Attention layers yield faithful explanations, but challenges persist for ultra-high-velocity streams, potentially requiring distributed adaptations. Future work includes expanding modality coverage and bolstering explanation robustness in imbalanced streams.

7. EXPERIMENTAL EVALUATION

This study presents a scalable, explainable classification framework for heterogeneous Big Data streams. By integrating multimodal autoencoding, attention-based interpretability, and adaptive online learning, it advances transparent AI for real-time analytics. The framework fosters compliance and user confidence in sensitive domains. Future extensions involve federated deployments for privacy-sensitive settings and refined explanation interfaces.

References

1. C. C. Aggarwal, Data Streams: Models and Algorithms. New York, NY, USA: Springer, 2007.
2. P. Domingos and G. Hulten, “Mining high-speed data streams,” in Proc. 6th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Boston, MA, USA, 2000, pp. 71–80.
3. H. M. Gomes et al., “Adaptive random forests for evolving data stream classification,” Mach. Learn., vol. 106, nos. 9–10, pp. 1469–1495, Oct. 2017.
4. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 1135–1144.
5. S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 4765–4774.
6. A. Vaswani et al., "Attention Is All You Need," in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 5998–6008.
7. D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in Proc. 2nd Int. Conf. Learn. Representations, 2014.
8. J. Xiong, “Decision-making model based on CNN for big data classification,” IEEE Access, vol. 7, pp. 105396–105407, 2019.
9. N. Gunasekara, B. Pfahringer, H. M. Gomes, A. Bifet, and Y. S. Koh, “Recurrent concept drifts on data streams,” in Proc. 33rd Int. Joint Conf. Artif. Intell., Jeju, South Korea, 2024, pp. 7997–8005.
10. J. Gama et al., "A Survey on Concept Drift Adaptation," ACM Comput. Surv., vol. 46, no. 4, pp. 1–37, 2014.
11. L. Deng et al., "Deep Learning: Methods and Applications," Found. Trends Signal Process., vol. 7, nos. 3–4, pp. 197–387, 2014.
12. Y. Zhang et al., "Multimodal Deep Learning: A Survey," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 6, pp. 3384–3402, 2022.