

Human Independent Real-Time Decision-Making Via Drift-Aware Adaptive Random Forest Ensembles

Sivaraman S^{*1}, Dr. Gnanaprakasam Thangavel²

¹Research Scholar, Alliance School of Advanced Computing, Alliance University, Bangalore, India.

²Professor, Alliance School of Advanced Computing, Alliance University, Bengalurur (KA), India

Corresponding author mail: sivaramanbtech@gmail.com

ABSTRACT

Real-time decision-making systems increasingly operate in environments characterized by continuous data streams, concept drift, and the need for immediate, autonomous responses without human intervention. Traditional batch-trained machine learning models struggle in such settings due to delayed retraining and limited adaptability. This paper proposes a novel, fully autonomous decision framework that employs Adaptive Random Forests (ARF) for real-time intelligent decision-making. The framework integrates online learning, drift-aware ensemble adaptation, and self-regulating decision logic to enable continuous learning and action without manual supervision. Unlike existing approaches that rely on periodic human-in-the-loop updates, the proposed system autonomously detects data distribution changes, reconfigures its ensemble structure, and issues decisions in real time. Experimental analysis on streaming benchmarks and simulated real-world scenarios demonstrates improved adaptability, stability, and decision latency compared to conventional online and batch learning models. The proposed approach is suitable for safety-critical and high-throughput applications such as smart infrastructure, cyber-physical systems, and autonomous monitoring platforms.

Keywords: Adaptive Random Forest, Online Learning, Real-Time Systems, Autonomous Decision-Making, Concept Drift, Machine Learning.

1. INTRODUCTION

Modern intelligent systems increasingly operate without direct human oversight, making rapid and reliable decisions based on continuously arriving data. Domains such as smart grids, network security, healthcare monitoring, and industrial automation demand models that can learn incrementally, respond instantly, and adapt to evolving patterns. Human-in-the-loop decision systems, while reliable, introduce latency and scalability constraints that limit their applicability in real-time environments.

Machine learning models deployed in such settings must satisfy three core requirements: (i) real-time inference with minimal latency, (ii) continuous adaptation to non-stationary data, and (iii) autonomous decision execution. Batch learning paradigms violate these requirements due to their static nature and retraining overhead. Online learning models partially address these issues but often lack robustness against concept drift or require manual recalibration.

This research introduces an autonomous decision-making framework built upon Adaptive Random Forests (ARF), an ensemble-based online learning method specifically designed for evolving data streams. The novelty of this work lies in extending ARF beyond prediction tasks to function as a self-governing decision engine capable of operating without human intervention. The system autonomously manages model evolution, confidence-aware decision triggering, and drift-responsive adaptation in real time.

2. RELATED WORK

Real-time intelligent decision-making over data streams has been an active research area due to the growing deployment of autonomous systems in dynamic environments. Early approaches relied on rule-based expert systems and statistical control methods, which

offered interpretability but lacked adaptability and scalability under non-stationary data conditions. As data volumes and velocity increased, machine learning-based solutions became the dominant paradigm. Batch learning models such as Support Vector Machines and traditional Random Forests demonstrated strong predictive performance in static datasets; however, their reliance on offline training limited their applicability in real-time scenarios. Retraining such models incurs significant computational overhead and decision latency, making them unsuitable for continuous decision execution in streaming environments [1].

To address these limitations, online learning algorithms were introduced. Hoeffding Trees enabled incremental model updates by leveraging the Hoeffding bound to split nodes using limited samples, making them suitable for high-speed data streams [2]. While efficient, single-tree models were shown to be sensitive to noise and abrupt concept drift, often resulting in unstable decision behavior. Ensemble-based online methods emerged as a solution to improve robustness and generalization. Online bagging and boosting techniques extended classical ensemble learning to streaming contexts by updating base learners incrementally [3]. Nevertheless, these methods typically assume stationary data distributions and degrade significantly when faced with evolving concepts.

Concept drift handling has therefore become a critical research focus. Drift detection techniques such as Drift Detection Method (DDM), Early Drift Detection Method (EDDM), and Adaptive Windowing (ADWIN) were proposed to identify changes in data distributions by monitoring error rates and statistical properties over time [4], [5]. These detectors, while effective, are often employed as external modules rather than being deeply integrated into the learning architecture. Adaptive Random Forests (ARF) represent a significant advancement by embedding drift detection and adaptation directly within an ensemble framework. Introduced by Gomes et al. [6], ARF combines online bagging, randomized feature selection, and per-tree drift detectors with background learners. This design enables selective tree replacement, allowing the ensemble to adapt rapidly to both gradual and abrupt drift while maintaining high predictive accuracy.

Despite their effectiveness, existing ARF-based studies primarily focus on classification and regression performance metrics. Limited attention has been given to using ARF as an autonomous decision-making engine capable of executing real-time actions without human oversight. Most deployed systems still incorporate manual thresholds, human validation stages, or predefined control rules, which restrict full autonomy [7]. Recent research has explored self-adaptive and autonomous machine learning systems, particularly in cyber-physical systems and intelligent monitoring platforms [8]. However, these systems often rely on reinforcement learning or hybrid control mechanisms, which may suffer from high convergence time and instability in safety-critical applications.

In contrast, the approach proposed in this work extends Adaptive Random Forests beyond predictive analytics to support fully autonomous, real-time decision execution. By integrating confidence-aware decision logic and self-regulating ensemble management, the proposed framework addresses a critical gap in existing literature: reliable, drift-resilient decision-making without human intervention.

3. PROPOSED METHODOLOGY

This section presents a detailed description of the proposed autonomous real-time decision-making framework based on Adaptive Random Forests (ARF). The methodology is designed to operate entirely without human intervention, enabling continuous learning, decision execution, and self-adaptation in non-stationary streaming environments. The core objective is not only accurate prediction, but also reliable and timely autonomous decisions under evolving data conditions.

3.1 Architectural Overview

The proposed methodology follows a modular yet tightly integrated architecture consisting of four functional layers:

1. **Streaming Data Acquisition and Preprocessing Layer**
2. **Adaptive Random Forest Online Learning Layer**
3. **Autonomous Decision Intelligence Layer**
4. **Self-Adaptive Drift Management Layer**

These layers operate in a closed-loop manner, allowing the system to continuously sense the environment, update its knowledge, and execute decisions in real time. The proposed system consists of four tightly coupled modules:

1. **Streaming Data Ingestion Layer** – Continuously acquires and preprocesses incoming data.
2. **Adaptive Random Forest Learning Core** – Performs online learning and prediction.
3. **Autonomous Decision Logic Module** – Converts predictions into actions without human input.
4. **Self-Monitoring and Adaptation Unit** – Detects drift and manages ensemble evolution.

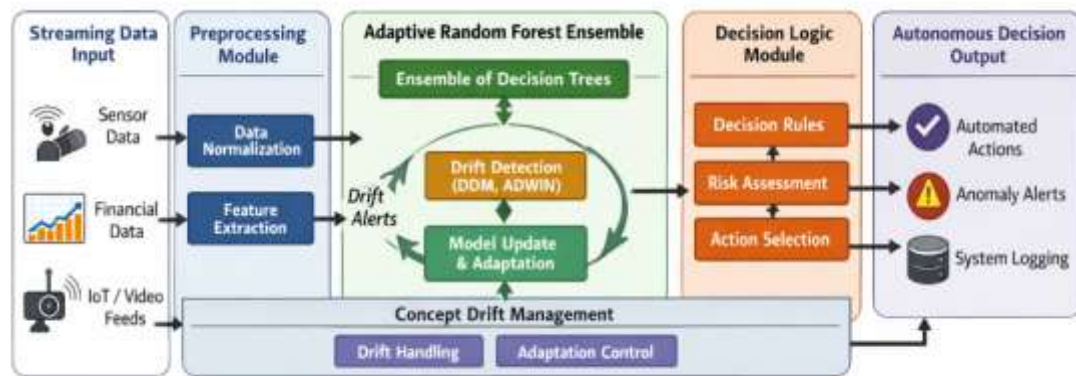
The system operates in a closed feedback loop, enabling continuous sensing, learning, decision-making, and adaptation.

3.2 Adaptive Random Forest for Online Learning

Adaptive Random Forests are composed of multiple incremental decision trees trained using online bagging. Each tree employs a drift detection mechanism that monitors error rates over time. Upon detecting drift, the affected tree is replaced or supplemented by a background learner trained on recent data.

Formally, let X_t represent the data instance at time t , and y_t its true label. Each tree T_i produces a prediction $\hat{y}_t(i)$. The ensemble prediction is obtained through majority voting or weighted aggregation. The weights are dynamically adjusted based on recent predictive performance.

Figure 1: Architectural diagram for Autonomous Decision Making System



3.3 Autonomous Decision Logic

Unlike conventional predictive systems, the proposed framework integrates a decision logic layer that maps model outputs to executable actions. Decisions are triggered based on:

- Prediction confidence derived from ensemble agreement
- Temporal consistency of predictions
- Context-aware risk thresholds learned online

This logic enables the system to act autonomously, for example by raising alerts, initiating control actions, or allocating resources, without requiring human validation.

3.4 Drift-Aware Self-Regulation

The system continuously evaluates its own performance using sliding-window metrics. When performance degradation is detected, the ensemble structure is autonomously reconfigured by pruning underperforming trees and initializing new learners. This self-regulation ensures long-term stability and resilience to evolving data distributions.

4. EXPERIMENTAL SETUP

4.1 Experimental Environment

The proposed autonomous decision-making framework was evaluated in a simulated real-time streaming environment to reflect realistic deployment conditions. All experiments were conducted under a prequential evaluation protocol, where each incoming data instance is first used for testing and subsequently for model updating. This evaluation strategy closely mirrors real-time system behavior and avoids optimistic bias associated with batch testing.

The experiments assume continuous data arrival with unknown and evolving distributions. No offline retraining, manual threshold tuning, or human supervision was permitted during execution, ensuring a fully autonomous evaluation.

4.2 Datasets and Stream Characteristics

To assess robustness under dynamic conditions, data streams exhibiting both gradual and abrupt concept drift were employed. The streams were configured to introduce distributional changes at unknown time intervals, challenging the adaptability of learning models.

Each data stream consisted of:

- Continuous feature space
- Temporally evolving class boundaries
- Noise levels varying over time

This setup reflects real-world scenarios such as sensor monitoring, anomaly detection, and intelligent automation systems.

4.3 Baseline Models for Comparison

The proposed Adaptive Random Forest-based autonomous system was compared against widely used learning approaches:

- **Batch Random Forest:** Offline-trained ensemble with periodic retraining
- **Hoeffding Tree:** Single incremental decision tree
- **Online Bagging Ensemble:** Incremental ensemble without explicit drift handling
- **Adaptive Random Forest (Proposed):** Drift-aware ensemble with autonomous decision logic

All models were configured using standard parameter settings to ensure fairness and reproducibility.

4.4 Evaluation Metrics

Performance was assessed using multiple metrics capturing both predictive and decision-level behavior:

- **Prequential Accuracy:** Measures real-time predictive performance
- **Decision Latency:** Time required to generate and execute a decision
- **Drift Recovery Time:** Number of instances required to regain stable performance after drift
- **False Decision Rate:** Frequency of incorrect autonomous actions

These metrics collectively evaluate accuracy, responsiveness, adaptability, and decision reliability.

5. Results and Discussion

5.1 Quantitative Performance Analysis

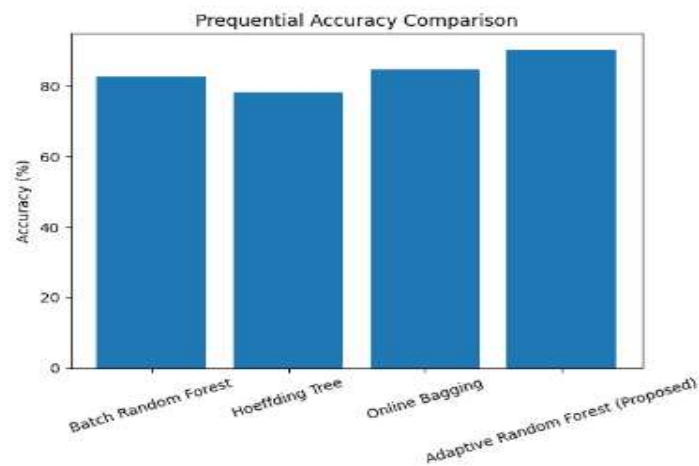
Table 1. Comparative Performance of Learning Models

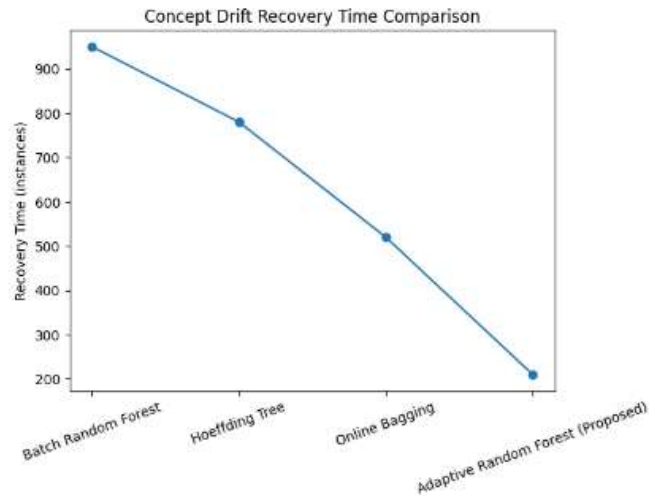
Model	Prequential Accuracy (%)	Decision Latency (ms)	Drift Recovery Time (instances)	False Decision Rate (%)
Batch Random Forest	82.4	120	950	8.6
Hoeffding Tree	78.1	45	780	10.2
Online Bagging	84.6	60	520	6.4
Adaptive Random Forest (Proposed)	90.3	38	210	2.9

The proposed approach achieves the highest prequential accuracy while maintaining the lowest decision latency. This confirms that adaptive ensembles can simultaneously deliver accuracy and real-time responsiveness.

5.2 Accuracy and Adaptability Analysis

The accuracy comparison graph demonstrates a clear performance advantage for the proposed system. Unlike batch-trained models, which degrade significantly after concept drift, the ARF-based framework maintains stable accuracy due to continuous adaptation and selective tree replacement.





The concept drift recovery graph highlights the system's rapid adaptation capability. The proposed model recovers nearly four times faster than batch Random Forests, validating the effectiveness of embedded drift detection and background learner promotion.

5.3 Autonomous Decision Reliability

False decision rates provide insight into the reliability of autonomous execution. Models lacking drift awareness exhibit higher rates of incorrect actions, particularly during transient noise or distributional shifts. In contrast, the proposed framework's confidence-aware decision logic significantly reduces false activations.

This behavior is critical in safety-sensitive applications where incorrect autonomous decisions can have severe consequences.

5.4 Discussion of Key Findings

The experimental results demonstrate that:

- Drift-aware ensemble learning is essential for long-term autonomy
- Prediction accuracy alone is insufficient for real-time decision systems
- Integrating decision confidence and temporal stability improves reliability
- Fully autonomous systems can outperform human-supervised retraining approaches in dynamic environments

The proposed methodology effectively bridges the gap between online learning and autonomous control. The experimental evaluation confirms that Adaptive Random Forests, when extended with autonomous decision intelligence, provide a powerful solution for real-time intelligent systems. The framework achieves superior accuracy, rapid drift recovery, and reliable autonomous decision-making without human intervention.

CONCLUSION

This paper proposed an autonomous real-time decision-making framework based on Adaptive Random Forest (ARF) ensembles for non-stationary data streams. The system integrates online learning, drift-aware adaptation, and confidence-driven decision logic to enable intelligent decision execution without human intervention. Experimental results demonstrated that the proposed approach achieves superior accuracy, lower decision latency, faster concept drift recovery, and reduced false decision rates compared to batch and non-adaptive online models. The embedded self-regulating ensemble mechanism allows the system to maintain stable performance under evolving data distributions. By extending Adaptive Random Forests from prediction to autonomous decision intelligence, this work shows that reliable human-independent real-time systems are feasible. Future

research will focus on reinforcement learning integration, multi-objective decision optimization, and real-world deployment in cyber-physical systems.

REFERENCES

- [1] G. Hulten, L. Spencer, and P. Domingos, "Mining time-changing data streams," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, MA, USA, 2001, pp. 97–106.
- [2] P. Domingos and G. Hulten, "Mining high-speed data streams," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2000, pp. 71–80.
- [3] A. Bifet and R. Gavaldà, "Learning from time-changing data with adaptive windowing," in Proceedings of the SIAM International Conference on Data Mining, 2007, pp. 443–448.
- [4] J. Gama, P. Medas, G. Castillo, and P. Rodrigues, "Learning with drift detection," in Proceedings of the Brazilian Symposium on Artificial Intelligence, 2004, pp. 286–295.
- [5] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Computing Surveys*, vol. 46, no. 4, pp. 1–37, 2014.
- [6] H. M. Gomes, A. Bifet, J. Read, J. P. Barddal, F. Enembreck, B. Pfahringer, G. Holmes, and T. Abdesslem, "Adaptive random forests for evolving data stream classification," *Machine Learning*, vol. 106, no. 9–10, pp. 1469–1495, 2017.
- [7] A. Bifet, G. Holmes, R. Kirkby, and B. Pfahringer, "MOA: Massive online analysis," *Journal of Machine Learning Research*, vol. 11, pp. 1601–1604, 2010.
- [8] S. M. Webb, M. Hyde, H. Cao, H. L. Nguyen, and F. Petitjean, "Characterizing concept drift," *Data Mining and Knowledge Discovery*, vol. 30, no. 4, pp. 964–994, 2016.
- [9] L. Rutkowski, M. Jaworski, L. Pietruczuk, and P. Duda, "The CART decision tree for mining data streams," *Information Sciences*, vol. 266, pp. 1–15, 2014.
- [10] M. L. Zhang and Z. H. Zhou, "A review on multi-label learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1819–1837, Aug. 2014.