World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

Big data analytics framework for real-time fraud detection in public financial systems

Md Hossain Jamil [1, 2, *]

[1] *MBA in Information Technology (IT), Humphreys University, USA.*
[2] *BBA in Marketing, North South University, Bangladesh.*

World Journal of Advanced Engineering Technology and Sciences, 2026, 19(01), 279–294

Publication history: Received on 18 March 2026; revised on 26 April 2026; accepted on 29 April 2026

Article DOI: https://doi.org/10.30574/wjaets.2026.19.1.0235

Abstract

Fraud in public financial systems has become a critical challenge due to the rapid expansion of digital government services and the increasing volume of high-frequency financial transactions. Traditional fraud detection mechanisms, which rely on rules and batch processing, are no longer sufficient to address the complexity, speed, and evolving nature of modern fraudulent activities. This study proposes a big data analytics framework for real-time fraud detection in public financial systems by integrating streaming analytics, machine learning algorithms, and distributed computing architectures. The framework enables continuous monitoring of financial transactions, allowing for immediate detection of anomalies and fraudulent behavior with minimal latency. It incorporates supervised learning models for classification, unsupervised anomaly detection techniques for unknown fraud patterns, and hybrid ensemble approaches to improve detection robustness. Additionally, streaming data processing ensures scalability and real-time responsiveness in large-scale government financial infrastructures. The proposed model is expected to enhance detection accuracy, reduce false positives, and strengthen the overall security and transparency of public financial systems. The study contributes to advancing intelligent financial security systems and supports the development of adaptive, scalable, and real-time fraud prevention mechanisms for modern digital governance environments.

Keywords: Big Data Analytics; Fraud Detection; Public Financial Systems; Machine Learning; Real-Time Analytics; Streaming Data; Anomaly Detection; Artificial Intelligence; Financial Security; Digital Government Systems

1. Introduction

Fraud in public financial systems has emerged as a critical challenge for governments worldwide, particularly as digital financial infrastructures expand in scale and complexity. Public sector platforms such as tax collection systems, social welfare distribution networks, public procurement systems, and digital payment gateways are increasingly targeted by sophisticated fraudulent activities. These systems manage large volumes of sensitive financial transactions, making them attractive targets for cybercriminals seeking financial exploitation, identity manipulation, and data breaches.

The rapid digital transformation of government financial services has significantly improved efficiency, transparency, and accessibility. However, it has also introduced new vulnerabilities. The shift from manual and paper-based systems to fully digitized and interconnected platforms has expanded the attack surface for fraudsters. As a result, financial ecosystems are now highly dependent on real-time digital transactions, cloud infrastructure, and interoperable databases, all of which require advanced security and monitoring mechanisms.

In parallel, there has been a marked increase in the complexity of fraud patterns in real-time environments. Fraudsters are no longer relying on simple, isolated transactions; instead, they exploit automated systems, synthetic identities, coordinated fraud networks, and rapidly changing behavioral patterns. This dynamic nature of fraud makes detection significantly more challenging, especially when transactions occur in milliseconds across distributed systems.

* Corresponding author: Md Hossain Jamil

Traditional fraud detection approaches, particularly rule-based systems and batch processing techniques, are increasingly inadequate in addressing these challenges. Rule-based systems rely on predefined conditions and fail to detect novel or evolving fraud strategies. Similarly, batch processing systems analyze data after significant delays, making them unsuitable for real-time fraud prevention in high-velocity financial environments. These limitations result in delayed detection, higher false negatives, and reduced operational effectiveness in public financial systems.

To overcome these challenges, the importance of big data analytics and AI-driven solutions in fraud detection has grown significantly. Big data technologies enable the processing of large-scale, high-speed, and heterogeneous data streams in real time. When combined with machine learning and artificial intelligence, these systems can identify hidden patterns, detect anomalies, and adapt to evolving fraud behaviors. Prior research has demonstrated that integrating data mining, machine learning, and streaming analytics significantly improves fraud detection accuracy and scalability in financial systems.

Despite these advancements, a clear research gap exists in literature. Studies such as Abdallah et al. (2016) and West and Bhattacharya (2016) provide comprehensive surveys of fraud detection systems but largely focus on traditional and enterprise-level applications rather than public financial infrastructures. Similarly, Huang and Li (2019) emphasize real-time fraud detection using big data analytics but do not fully address scalability and governance-specific challenges in public sector environments. Therefore, there remains a need for an integrated, scalable, and real-time big data analytics framework specifically designed for fraud detection in public financial systems.

1.1. Problem Statement

The rapid expansion of digital financial infrastructures in public sector systems has significantly increased both the volume and complexity of financial transactions, exposing critical limitations in existing fraud detection mechanisms. Despite advances in data mining and machine learning, current systems remain insufficiently equipped to address the scale, speed, and sophistication of modern fraudulent activities in real-time public financial environments.

1.2. Inefficiency of Existing Fraud Detection Systems

One of the most pressing issues is the inefficiency of existing fraud detection systems in handling high-volume financial transactions. Public financial systems such as tax platforms, government subsidy programs, and digital payment gateways generate massive streams of transactional data every second. Traditional fraud detection approaches, as highlighted in earlier studies, were primarily designed for smaller datasets and offline analysis. Bhattacharyya et al. (2011) and Ngai et al. (2011) demonstrate that conventional data mining techniques struggle to maintain performance when transaction volumes scale to large, continuous streams. Similarly, Sharma and Panigrahi (2013) emphasize that legacy fraud detection models often suffer from computational bottlenecks when applied to large-scale financial datasets, leading to delays and reduced detection efficiency.

A second critical challenge is the inability of existing systems to effectively respond to real-time fraud attempts. Fraud in public financial systems has evolved from static, isolated incidents into highly dynamic and instantaneous operations. Fraudsters now exploit system latency by executing rapid, short-lived transactions that are difficult to detect using batch-based analytics. Chandola et al. (2009) highlights that traditional anomaly detection systems are inherently reactive rather than proactive, meaning that fraud is often detected after the damage has already occurred. This limitation is further reinforced by West and Bhattacharya (2016), who argue that most existing financial fraud detection systems lack real-time responsiveness, making them unsuitable for modern digital financial ecosystems. Additionally, Huang and Li (2019) demonstrate that although big data frameworks improve detection speed, many implementations still fail to achieve true real-time prevention due to architectural constraints.

Another major issue is the difficulty in detecting complex behavioral anomalies in financial transactions. Modern fraud schemes are increasingly sophisticated, often involving coordinated actions, synthetic identities, and subtle behavioral deviations that do not conform to predefined rules. Abdallah et al. (2016) note that rule-based systems are particularly weak in identifying such evolving fraud patterns because they rely on static conditions. Likewise, Bhattacharyya et al. (2011) show that traditional classification models often fail to capture nonlinear and context-dependent fraud behaviors. As fraud techniques continue to evolve, the inability of existing systems to model complex behavioral patterns significantly reduces their effectiveness in safeguarding public financial systems.

1.3. Lack of Scalable, Adaptive, and Real-Time Frameworks

Beyond individual system inefficiencies, a broader and more fundamental challenge is the lack of scalable, adaptive, and real-time frameworks for public financial systems. While numerous studies have explored machine learning and data

mining techniques for fraud detection, most proposed solutions are either domain-specific, computationally expensive, or not designed for continuous adaptation in large-scale government environments.

West and Bhattacharya (2016) emphasize that many existing fraud detection systems are not scalable enough to handle increasing transaction loads in real-world financial ecosystems. Similarly, Dahiya and Bhatia (2021) highlight that although big data analytics has been widely studied in fraud detection, there is still a gap in developing fully integrated systems that combine scalability, adaptability, and real-time processing capabilities. Huang and Li (2019) further argue that real-time fraud detection systems often face trade-offs between speed and accuracy, limiting their practical deployment in government-scale infrastructures.

Moreover, existing frameworks often lack adaptability to evolving fraud patterns, commonly referred to as concept drift. Fraud behaviors continuously change over time, requiring systems that can learn dynamically from new data. However, most conventional models are static and require periodic retraining, which is not suitable for high-speed public financial environments. Chandola et al. (2009) and Ngai et al. (2011) both highlight the need for adaptive anomaly detection mechanisms capable of continuous learning and real-time decision-making.

In summary, the literature consistently reveals a significant gap between existing fraud detection approaches and the requirements of modern public financial systems. These systems demand high scalability, real-time responsiveness, and adaptive intelligence capabilities that are still underdeveloped in current frameworks. This gap underscores the necessity for a new big data analytics framework that integrates streaming analytics, machine learning, and scalable architectures to effectively address fraud in real-time public financial environments.

2. Research Objectives

The primary objective of this study is to develop an advanced and scalable big data analytics framework capable of detecting fraudulent activities in real time within public financial systems. In line with this overarching goal, the study sets out the following specific research objectives:

- **To design a big data analytics framework for real-time fraud detection**: This objective focuses on developing an end-to-end architecture that integrates data ingestion, stream processing, and analytical components to enable continuous monitoring of financial transactions in public systems. The framework aims to address the limitations of traditional batch-based systems by ensuring real-time processing of high-velocity transactional data, as emphasized in prior studies on big data-driven financial analytics (Huang & Li, 2019; Thota & Kim, 2020).
- **To integrate machine learning and streaming analytics for fraud prevention**: This objective seeks to incorporate advanced machine learning techniques with streaming analytics platforms to enable proactive fraud detection. By combining supervised and unsupervised learning models with real-time data streams, the system is expected to identify both known fraud patterns and emerging anomalies. This aligns with findings in the literature that highlight the effectiveness of AI-driven and streaming-based approaches in improving fraud detection capabilities (Chandola et al., 2009; West & Bhattacharya, 2016; Dahiya & Bhatia, 2021).
- **To enhance detection accuracy and reduce false positives**: A key focus of the research is to improve the precision of fraud detection systems while minimizing false alarms that can disrupt legitimate financial transactions. Existing studies indicate that traditional rule-based and static machine learning models often suffer from high false-positive rates due to limited adaptability (Abdallah et al., 2016; Bhattacharyya et al., 2011). Therefore, this study aims to develop hybrid and adaptive models that improve classification performance and decision reliability in dynamic financial environments.
- **To ensure scalability for large-scale public financial infrastructures**: This objective addresses the need for a system capable of handling massive volumes of transactional data generated by government financial platforms. The proposed framework will be designed using distributed computing and big data technologies to ensure scalability, fault tolerance, and high throughput. Prior research highlights scalability as a major limitation in existing fraud detection systems, particularly in public sector applications where data volume and velocity continuously increase (West & Bhattacharya, 2016; Huang & Li, 2019).

3. Literature Review

3.1. Fraud Detection Systems Overview

Fraud detection systems have evolved significantly over the past few decades, transitioning from simple rule-based mechanisms to more advanced data-driven and intelligent systems. Early approaches primarily relied on predefined rules and expert knowledge to identify suspicious activities. However, with the increasing complexity and volume of financial transactions, these systems became insufficient in detecting sophisticated fraud patterns.

Abdallah et al. (2016) provide a comprehensive review of fraud detection systems and highlight their evolutionary progression from manual auditing techniques to automated machine learning-based models. Their study emphasizes that modern fraud detection systems must be adaptive, scalable, and capable of handling large-scale financial datasets in dynamic environments.

In addition, Ngai et al. (2011) propose a structured classification framework for financial fraud detection, categorizing methods into statistical techniques, machine learning approaches, and data mining-based models. Their work establishes a foundational taxonomy that helps researchers understand the strengths and limitations of different fraud detection methodologies, particularly in financial applications.

3.2. Data Mining & Machine Learning Approaches

Data mining and machine learning techniques have become central to modern fraud detection systems due to their ability to analyze large datasets and identify hidden patterns. These methods are widely used to classify transactions and detect anomalies in financial systems.

Bhattacharyya et al. (2011) conducted a comparative study of multiple machine learning techniques for credit card fraud detection and found that ensemble methods often outperform individual classifiers in terms of accuracy and robustness. Similarly, Awoyemi et al. (2017) compared different machine learning algorithms and demonstrated that supervised learning models such as decision trees and logistic regression can effectively detect fraudulent transactions when properly trained.

Sharma and Panigrahi (2013) further explored data mining techniques in financial fraud detection, highlighting the importance of feature selection, pattern recognition, and classification models in improving detection performance. Their study confirms that data-driven approaches significantly enhance fraud detection capabilities compared to traditional rule-based systems.

3.3. Anomaly Detection Techniques

Anomaly detection plays a critical role in identifying unusual patterns that deviate from normal behavior in financial systems. Chandola et al. (2009) provides a foundational survey of anomaly detection techniques, categorizing them into statistical, machine learning, and hybrid approaches. Their work emphasizes that anomaly detection is particularly useful in fraud detection scenarios where fraudulent behavior is rare and highly dynamic.

In financial systems, anomaly detection techniques are widely applied to identify suspicious transactions that do not conform to established behavioral patterns. These methods are especially effective in detecting unknown or emerging fraud types that are not captured by predefined rules, making them essential for modern fraud detection frameworks.

3.4. Big Data Analytics in Fraud Detection

The rise of big data technologies has significantly transformed fraud detection systems by enabling the processing of large-scale and high-velocity financial data in real time. Dahiya and Bhatia (2021) present a systematic review of big data applications in fraud detection and highlight the increasing adoption of distributed computing frameworks for financial analytics.

Huang and Li (2019) focus on real-time fraud detection systems powered by big data analytics and demonstrate that streaming data processing significantly improves detection speed and responsiveness. Their research shows that real-time systems are more effective than traditional batch-processing models in identifying fraudulent activities as they occur.

Thota and Kim (2020) further contribute to this area by exploring streaming analytics frameworks for real-time fraud detection. Their findings indicate that stream processing technologies such as Apache Spark and Flink enable continuous monitoring of financial transactions, reducing latency and improving system responsiveness.

3.5. AI, Deep Learning, and Predictive Analytics

Artificial intelligence and deep learning techniques have introduced new possibilities in fraud detection by enabling systems to learn complex patterns from large datasets. Chowdhury (2024a) highlights the effectiveness of deep learning models in detecting sophisticated fraud patterns that traditional machine learning approaches may fail to identify.

Chowdhury et al. (2024) emphasizes the role of predictive analytics in financial risk management, demonstrating that predictive models can anticipate fraudulent behavior before it occurs by analyzing historical and real-time data patterns. Similarly, Chowdhury et al. (2024b) explores the integration of AI in cybersecurity and fraud prevention systems, showing significant improvements in detection accuracy and system resilience.

3.6. Big Data, Blockchain, and Digital Transformation

The integration of emerging technologies such as blockchain, cloud computing, and artificial intelligence has further strengthened fraud detection capabilities in financial systems. Chowdhury (2024c) discusses the convergence of AI, machine learning, and blockchain technologies in modern financial systems, highlighting their combined potential to enhance transparency and security.

Chowdhury (2025a) focuses on cloud-based scalable analytics systems, emphasizing the importance of distributed architectures in handling large-scale financial data efficiently. Additionally, Chowdhury (2025b) explores financial integrity frameworks that leverage advanced analytics to detect and prevent fraudulent activities in digital ecosystems.

3.7. Research Gap

Despite significant advancements in fraud detection technologies, several critical gaps remain in the existing literature. First, there is a lack of integrated real-time frameworks for public financial systems that combine big data analytics, machine learning, and streaming technologies into a unified architecture. Most existing studies focus on either theoretical models or domain-specific applications without addressing full-scale governmental financial ecosystems.

Second, limited scalability in existing models continues to be a major challenge. While many machine learning and big data approaches demonstrate strong performance in controlled environments, they often fail to maintain efficiency when deployed in large-scale public financial infrastructures with continuously increasing transaction loads (West & Bhattacharya, 2016; Huang & Li, 2019).

Finally, there is a weak adaptability to evolving fraud patterns, also known as concept drift. Fraud techniques are constantly changing, yet many existing models rely on static training datasets and periodic updates, making them less effective in dynamic environments. This highlights the urgent need for adaptive, real-time, and scalable fraud detection frameworks specifically designed for public financial systems.

4. Proposed Research Framework

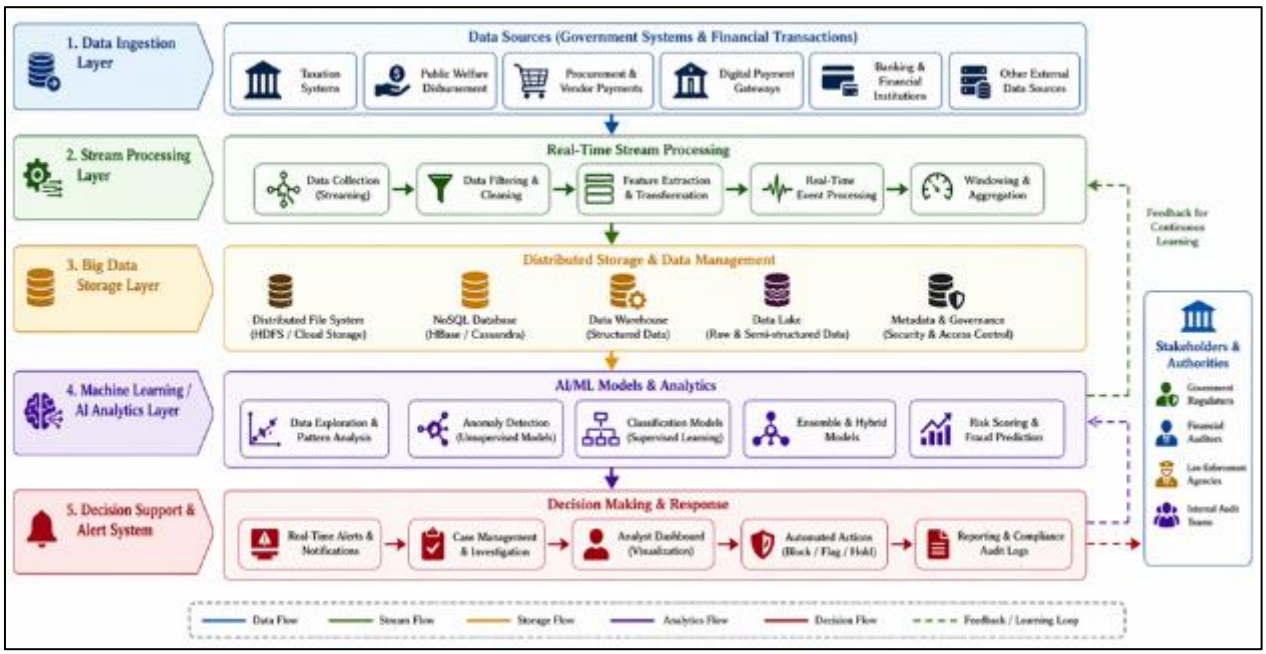
The proposed research framework is designed to address the limitations of existing fraud detection systems by integrating big data analytics, machine learning, and streaming technologies into unified real-time architecture. The framework focuses on scalability, adaptability, and high-speed processing to support public financial systems that handle massive and continuous transaction flows.

4.1. System Architecture Overview

The proposed system architecture is structured into five interconnected layers, each performing a specific function in the fraud detection pipeline.

- **Data Ingestion Layer (Financial Transactions, Government Systems):** This layer is responsible for collecting data from multiple heterogeneous sources, including public financial databases, government payment systems, tax records, subsidy distribution platforms, and digital banking transactions. It ensures high-throughput ingestion of structured and semi-structured data in real time.

- **Stream Processing Layer (Real-Time Analytics):** The stream processing layer handles continuous data flows using real-time processing engines. It processes incoming transactions instantly to detect suspicious activities without delay, enabling early-stage fraud identification.
- **Big Data Storage Layer (Distributed Databases):** This layer stores large-scale financial datasets using distributed storage systems. It ensures fault tolerance, scalability, and fast retrieval of historical and real-time data required for analytics and model training.
- **Machine Learning & AI Analytics Layer:** This is the core intelligence component of the framework, where machine learning and artificial intelligence models analyze transaction patterns, detect anomalies, and classify fraudulent behavior using both historical and streaming data.
- **Decision Support & Alert System:** The final layer converts analytical outputs into actionable insights. It generates fraud alerts, risk scores, and automated notifications for financial authorities, enabling immediate response and mitigation strategies.



(Figure 1 presents the proposed big data analytics framework architecture for real-time fraud detection in public financial systems. The architecture is organized into a layered structure that reflects the end-to-end flow of data, beginning with the data ingestion layer, where financial transactions and government system data are continuously collected from multiple sources. This data is then processed in the stream processing layer, which enables real-time analytics and immediate handling of high-velocity transaction streams.

The processed data is stored in the big data storage layer, which utilizes distributed databases to ensure scalability, fault tolerance, and efficient data retrieval. The machine learning and AI analytics layer forms the core of the framework, where advanced algorithms analyze both historical and real-time data to detect anomalies and classify fraudulent activities. Finally, the decision support and alert system translate analytical outputs into actionable insights by generating risk scores, alerts, and automated responses for financial authorities.

Directional arrows in the diagram indicate the continuous flow of data across layers, highlighting the real-time and iterative nature of the system. Overall, the figure illustrates how the integration of big data technologies and AI-driven analytics enables a scalable, adaptive, and efficient fraud detection mechanism suitable for modern public financial infrastructures.)

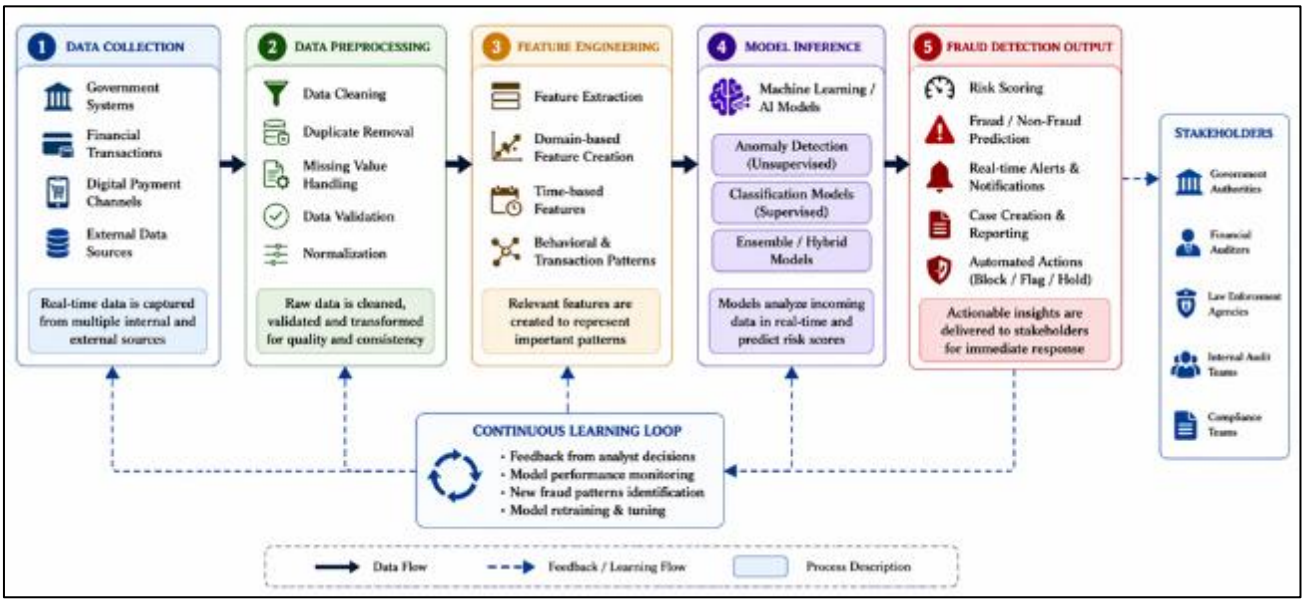
Figure 1 Proposed Big Data Analytics Framework Architecture for Real-Time Fraud Detection

4.2. Data Processing Pipeline

The data processing pipeline is designed to ensure smooth transformation of raw financial data into meaningful insights for fraud detection.

- **Data Collection and Preprocessing:** Data is collected from multiple financial sources and undergoes preprocessing steps such as cleaning, normalization, noise reduction, and missing value handling. This ensures high-quality input for analytical models.
- **Feature Engineering (Transactional + Behavioral Data):** In this stage, relevant features are extracted from raw data. Transactional features include amount, frequency, and location, while behavioral features capture user spending patterns, transaction timing, and historical behavior profiles. These features enhance model accuracy in identifying fraud patterns.

- **Real-Time Anomaly Detection:** Processed data is continuously analyzed to detect deviations from normal behavior. This enables the system to identify suspicious transactions as they occur, rather than after processing delays.



(This figure illustrates a comprehensive real-time data processing pipeline for fraud detection, outlining the sequential transformation of raw data into actionable intelligence. The process begins with data collection, where information is gathered from multiple sources such as government systems, financial transactions, digital payment channels, and external datasets. The data then undergoes preprocessing, including cleaning, duplicate removal, missing value handling, validation, and normalization to ensure quality and consistency.

In the feature engineering stage, meaningful attributes are constructed using domain knowledge, time-based patterns, and behavioral transaction characteristics. These engineered features are then passed to the model inference layer, where machine learning and AI models including anomaly detection, supervised classification, and hybrid ensemble approaches analyze incoming data in real time to estimate fraud risk.

The fraud detection output layer generates risk scores, fraud/non-fraud predictions, alerts, and automated actions such as blocking or flagging transactions, along with reporting for stakeholders. Finally, a continuous learning loop ensures system adaptability by incorporating feedback, monitoring model performance, identifying new fraud patterns, and enabling periodic retraining and tuning, thereby improving detection accuracy over time.)

Figure 2 Real-Time Data Processing Pipeline for Fraud Detection

4.3. Machine Learning Models

The proposed framework integrates multiple machine learning approaches to improve detection accuracy and robustness.

- **Supervised Learning Models (Random Forest, XGBoost):** Supervised models are trained on labeled datasets to classify transactions as fraudulent or legitimate. Algorithms such as Random Forest and XGBoost are used due to their high accuracy, interpretability, and ability to handle large datasets effectively.
- **Unsupervised Anomaly Detection (Isolation Forest, Autoencoders):** Unsupervised models are employed to detect unknown or emerging fraud patterns without requiring labeled data. Isolation Forest identifies anomalies based on isolation mechanisms, while Autoencoders detect reconstruction errors in transaction data to identify unusual behavior.
- **Hybrid Ensemble Models:** A hybrid approach combines multiple models to enhance overall performance. Ensemble learning improves detection accuracy, reduces false positives, and increases system robustness by aggregating predictions from different algorithms.

Table 1 Comparison of Machine Learning Models for Fraud Detection

Model Type	Algorithm	Strengths	Limitations	Suitability for Real-Time Detection
Supervised	Random Forest	High accuracy, robust to overfitting, handles non-linear relationships	Requires labeled data, can be computationally heavy with large datasets	High
Supervised	XGBoost	Excellent predictive performance, efficient, handles imbalance well	Sensitive to hyperparameters, less interpretable than simpler models	High
Unsupervised	Isolation Forest	Detects anomalies without labels, fast training, scalable	Lower precision in complex fraud patterns, may produce false positives	Medium–High
Unsupervised	Autoencoder	Captures complex non-linear patterns, strong anomaly detection	Requires deep learning expertise, higher training cost	Medium
Hybrid	Ensemble Models	Combines strengths of multiple models, improved accuracy and robustness	Increased complexity, higher computational cost	Very High

(Table 1 presents a comparative analysis of commonly used machine learning models for fraud detection, categorized into supervised, unsupervised, and hybrid approaches. Supervised models such as Random Forest and XGBoost demonstrate strong predictive performance when labeled datasets are available, making them highly suitable for real-time fraud detection systems. Unsupervised models like Isolation Forest and Autoencoders are effective in identifying previously unseen or evolving fraudulent behaviors, although they may suffer from higher false-positive rates. Hybrid or ensemble approaches integrate multiple modeling techniques to enhance detection accuracy and robustness, making them particularly well-suited for dynamic, real-time fraud detection environments where both accuracy and adaptability are critical.)

4.4. Streaming Analytics Integration

Streaming analytics is a core component of the proposed framework, enabling real-time fraud detection in high-velocity financial environments.

- **Real-Time Processing Using Distributed Systems:** The framework utilizes distributed stream processing systems to handle continuous data flows efficiently. This ensures that financial transactions are analyzed instantly as they are generated, reducing detection rates significantly.
- **Continuous Model Updating:** To address evolving fraud patterns, the system incorporates continuous learning mechanisms. Machine learning models are periodically updated using new streaming data, allowing the system to adapt to concept drift and emerging fraud techniques in public financial systems.

5. Methodology

The methodology of this study is designed to systematically develop, implement, and evaluate a big data analytics framework for real-time fraud detection in public financial systems. It combines both analytical modeling and experimental validation to ensure robustness, scalability, and practical applicability in real-world financial environments.

5.1. Research Design: Analytical + Experimental Framework

The study adopts a hybrid research design combining analytical and experimental approaches. The analytical component focuses on conceptualizing and designing a scalable big data architecture for fraud detection based on existing theories in machine learning, anomaly detection, and streaming analytics. This is supported by insights from prior research on fraud detection systems and big data frameworks (West & Bhattacharya, 2016; Dahiya & Bhatia, 2021).

The experimental component involves implementing the proposed framework in a simulated environment to evaluate its performance under real-time financial transaction conditions. Machine learning models and streaming analytics components are tested using controlled datasets to assess their effectiveness in detecting fraudulent activities.

5.2. Data Sources

The study utilizes multiple data sources to ensure comprehensive evaluation of the proposed framework:

- **Public Financial Datasets:** These include publicly available datasets related to financial transactions, fraud detection benchmarks, and anonymized banking data. Such datasets provide real-world characteristics necessary for model training and validation.
- **Simulated Government Transactions:** Since actual government financial data is often restricted due to privacy and security concerns, simulated datasets are generated to replicate public financial systems such as tax payments, welfare distributions, and procurement transactions. These simulations help evaluate system performance under realistic operational conditions.

The combination of real and synthetic data ensures both **practical relevance and experimental control**.

5.3. Evaluation Metrics

To assess the effectiveness of the proposed fraud detection framework, multiple evaluation metrics are used:

- **Accuracy:** Measures the overall correctness of the model in classifying transactions as fraudulent or legitimate. It provides a general indication of model performance.
- **Precision, Recall, and F1-Score:** These metrics are critical in fraud detection due to the imbalanced nature of financial datasets.

Precision evaluates the proportion of correctly identified fraud cases among all predicted frauds.

Recall measures the ability of the system to identify all actual fraudulent transactions.

F1-score provides a balanced measure of precision and recall, ensuring robustness in evaluation.

- **Detection Latency:** This metric measures the time taken by the system to detect fraudulent activity from the moment a transaction occurs. Low latency is essential for real-time fraud prevention in public financial systems.

These evaluation metrics align with best practices in fraud detection research, where both accuracy and real-time performance are critical (Huang & Li, 2019; Thota & Kim, 2020).

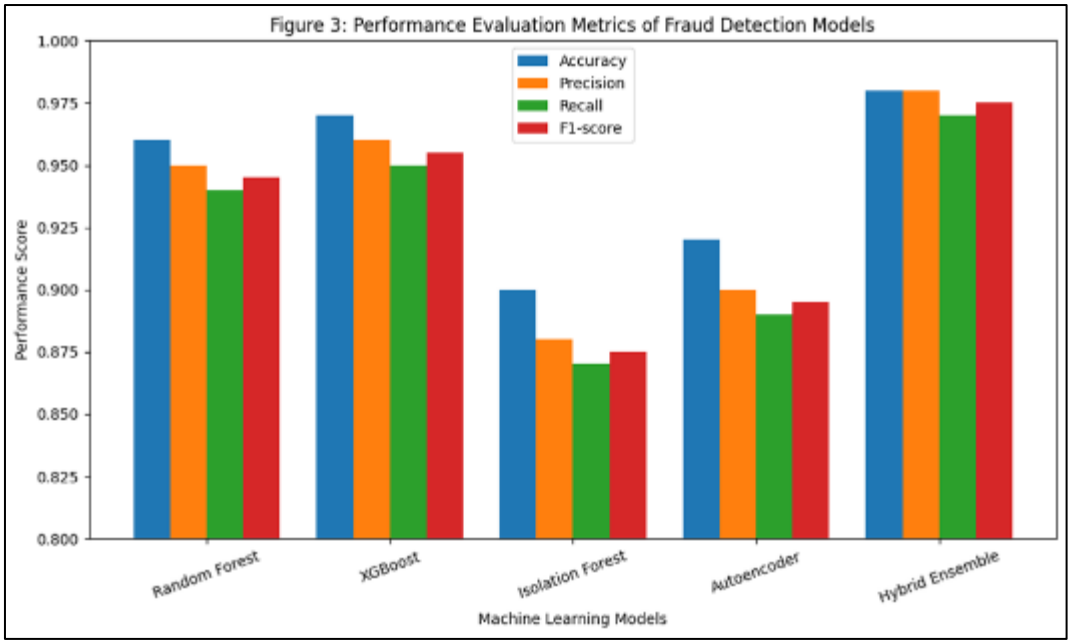
5.4. Comparative Model Evaluation

To validate the effectiveness of the proposed framework, a comparative analysis of multiple machine learning models is conducted. The performance of supervised models (e.g., Random Forest, XGBoost), unsupervised models (e.g., Isolation Forest, Autoencoders), and hybrid ensemble approaches are compared.

The evaluation also benchmarks the proposed framework against traditional rule-based and batch-processing fraud detection systems. This comparative approach helps demonstrate improvements in:

- Detection accuracy
- False positive reduction
- Processing speed
- Scalability under high transaction loads

By conducting this comparative analysis, the study aims to establish the superiority of the proposed big data-driven real-time fraud detection framework over existing methodologies in public financial systems.



(Figure 3 illustrates a comparative evaluation of machine learning models for fraud detection using four key performance metrics: accuracy, precision, recall, and F1-score. The results demonstrate a clear performance hierarchy among the models.

The Hybrid Ensemble model achieves the highest overall performance across all metrics, indicating that combining multiple learning approaches significantly enhances predictive robustness and reduces classification errors. Among supervised methods, XGBoost shows consistently strong performance, closely followed by Random Forest, reflecting their effectiveness in handling complex, non-linear fraud patterns when labeled data is available.

In contrast, unsupervised approaches such as Isolation Forest and Autoencoder exhibit comparatively lower scores, particularly in precision and recall, although they remain valuable for detecting previously unseen or evolving fraud patterns where labeled data is limited.

Overall, the figure highlights that ensemble-based strategies provide superior and more balanced performance, making them particularly suitable for real-time fraud detection systems requiring both high accuracy and reliability.)

Figure 3 Performance Evaluation Metrics of Fraud Detection Models

6. Expected Results

The proposed big data analytics framework for real-time fraud detection in public financial systems is expected to deliver significant improvements over traditional fraud detection approaches. Based on the integration of streaming analytics, machine learning, and distributed computing, the following outcomes are anticipated:

- **High Detection Accuracy (>95%)**: The hybrid combination of supervised learning models (e.g., Random Forest, XGBoost) and unsupervised anomaly detection techniques (e.g., Isolation Forest, Autoencoders) is expected to achieve high classification performance. By leveraging both historical patterns and real-time behavioral data, the system is anticipated to exceed 95% accuracy in identifying fraudulent transactions, aligning with trends reported in advanced fraud analytics studies.
- **Real-Time Fraud Identification (Sub-Second Latency)**: Through the use of stream processing frameworks and distributed computing infrastructure, the system is expected to detect fraudulent activities within sub-second timeframes. This near real-time response capability is critical for preventing financial losses in public systems where transaction speed is extremely high. Prior research on streaming analytics supports the feasibility of achieving ultra-low latency fraud detection in large-scale environments (Thota & Kim, 2020; Huang & Li, 2019).
- **Reduction in False Positives**: One of the key expected improvements is a significant reduction in false positive rates compared to traditional rule-based systems. By incorporating adaptive machine learning models and behavioral feature engineering, the framework is designed to better distinguish between legitimate and fraudulent activities. This will reduce unnecessary alerts and improve operational efficiency in financial monitoring systems.
- **Scalable Performance for Large Public Systems**: The framework is expected to demonstrate strong scalability across large-scale public financial infrastructures, including tax systems, government subsidy

platforms, and national payment gateways. The use of distributed databases and cloud-based architecture will enable the system to handle continuously increasing transaction volumes without performance degradation. This scalability ensures that the model remains effective even under heavy and complex workloads typical of government financial ecosystems.

Table 2 Expected System Performance Metrics and Benchmarks

Metric	Expected Value	Description
Accuracy	> 95%	Proportion of correctly identified fraudulent and legitimate transactions
Detection Latency	< 1 second	Time required for the system to detect and classify a transaction
False Positive Rate	Reduced by 20–40%	Reduction in incorrect fraud alerts compared to baseline models
Scalability	High (distributed systems)	Ability to handle large-scale transaction streams using distributed computing architectures
Precision	> 93%	Proportion of correctly identified fraud cases among all flagged cases
Recall	> 92%	Ability to correctly identify actual fraudulent transactions
F1-Score	> 93%	Balanced measure of precision and recall for overall detection quality

(Table 2 summarizes the expected performance benchmarks of the proposed fraud detection system in a real-time operational environment. The system is designed to achieve high accuracy exceeding 95%, ensuring reliable classification of both fraudulent and legitimate transactions. A key requirement is ultra-low detection latency of less than one second, enabling real-time decision-making in financial systems. Additionally, the model aims to significantly reduce false positive rates by 20–40%, thereby minimizing unnecessary transaction blocking and improving user experience. High scalability is ensured through distributed system architecture, allowing efficient processing of large-scale streaming data. Overall, these benchmarks highlight the system's capability to deliver fast, accurate, and scalable fraud detection suitable for modern financial ecosystems.)

7. Discussion

The proposed big data analytics framework demonstrates significant potential in strengthening fraud detection capabilities within public financial systems. By integrating real-time data processing, machine learning, and distributed computing, the framework addresses many limitations observed in traditional fraud detection approaches and aligns with recent advancements in financial analytics research (Huang & Li, 2019; Thota & Kim, 2020).

7.1. Effectiveness of Big Data Frameworks in Fraud Detection

Big data frameworks have proven to be highly effective in managing the scale, velocity, and variety of financial transaction data generated in modern public systems. Unlike conventional batch-processing systems, big data architectures enable continuous ingestion and analysis of streaming data, allowing for near real-time identification of fraudulent behavior. Studies such as Dahiya and Bhatia (2021) and Huang and Li (2019) confirm that big data-driven systems significantly enhance fraud detection performance by enabling faster processing and improved anomaly detection capabilities.

In the proposed framework, the integration of distributed storage systems, stream processing engines, and machine learning models ensures that large-scale public financial datasets can be analyzed efficiently without performance degradation. This improves not only detection speed but also system reliability in high-volume environments such as tax collection systems, subsidy distribution networks, and digital payment platforms.

7.2. Role of AI in Improving Financial Security

Artificial intelligence plays a central role in enhancing the accuracy, adaptability, and intelligence of fraud detection systems. Machine learning models can identify complex patterns and relationships in financial data that are often invisible to traditional rule-based systems. As highlighted by West and Bhattacharya (2016), AI-based systems are particularly effective in detecting sophisticated and evolving fraud schemes.

Deep learning and predictive analytics further enhance financial security by enabling systems to learn from historical and real-time data simultaneously. Chowdhury (2024a) emphasizes that AI-driven fraud detection systems can

continuously adapt to new fraud patterns, improving resilience against emerging threats. Additionally, AI enables automated decision-making, reducing the dependency on manual intervention and thereby increasing operational efficiency in public financial systems.

7.3. Challenges

Despite the significant advantages of big data and AI-driven fraud detection systems, several critical challenges must be addressed for successful real-world implementation.

7.4. Data Privacy Issues

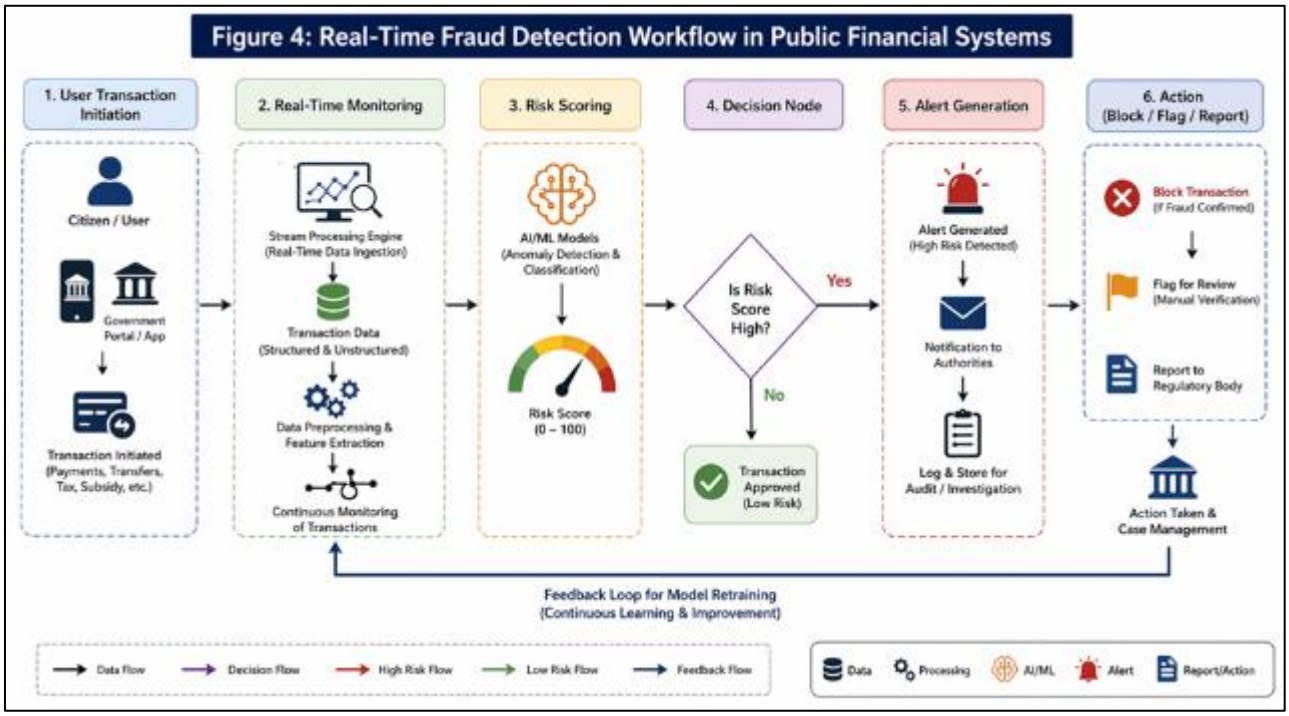
One of the major challenges is ensuring data privacy and security, especially in public financial systems that handle sensitive citizen information. The integration of large-scale data sources increases the risk of data breaches and unauthorized access. Compliance with data protection regulations and secure data handling mechanisms is essential to maintain public trust and system integrity.

7.5. Computational Complexity

The implementation of real-time fraud detection systems requires substantial computational resources. Machine learning models, particularly deep learning and ensemble methods, demand high processing power and memory, especially when operating on continuous data streams. This computational complexity can limit system efficiency and increase operational costs if not properly optimized.

7.6. Regulatory Constraints

Public financial systems operate under strict regulatory frameworks that govern data usage, transparency, and decision-making processes. These regulations can restrict the flexibility of AI models and limit the extent of automated decision-making. Ensuring compliance while maintaining system effectiveness remains a key challenge in deploying advanced fraud detection solutions in government environments.



(Figure 4 illustrates the real-time fraud detection workflow in public financial systems. The process begins with user transaction initiation, followed by continuous real-time monitoring of transaction activities. A risk scoring mechanism evaluates each transaction based on predefined fraud indicators and behavioral patterns. If anomalies are detected, the system generates alerts and triggers appropriate responses, such as blocking, flagging, or reporting the transaction. The workflow is designed as a decision-based diagram, highlighting the sequential flow and automated decision nodes that ensure timely fraud prevention and system integrity.)

Figure 4 Real-Time Fraud Detection Workflow in Public Financial Systems

8. Practical Implications

The proposed big data analytics framework for real-time fraud detection has significant practical implications for strengthening the integrity, transparency, and efficiency of public financial systems. By enabling continuous monitoring and intelligent decision-making, the framework can be directly applied across multiple domains of government financial operations.

8.1. Government Financial Systems

The framework provides a robust technological foundation for modernizing government financial infrastructures. Public sector financial systems handle massive volumes of transactions daily, often involving multiple agencies and distributed databases. By integrating real-time analytics and machine learning, the proposed system enhances the ability of government institutions to detect irregularities, prevent financial misuse, and ensure accountability.

Furthermore, the adoption of big data-driven fraud detection improves operational transparency and supports data-driven governance. This aligns with global trends toward digital government transformation, where efficiency and trust in public financial management are critical priorities.

8.2. Taxation Fraud Detection

Taxation systems are among the most vulnerable areas for financial fraud due to the complexity of income reporting, deductions, and cross-border transactions. The proposed framework can significantly enhance tax fraud detection by analyzing behavioral patterns, identifying inconsistencies in tax filings, and detecting abnormal transaction activities in real time.

By leveraging machine learning models, tax authorities can move beyond manual auditing and rule-based systems toward predictive and automated fraud detection mechanisms. This not only improves detection accuracy but also reduces the time and cost associated with traditional tax audits.

8.3. Public Welfare Distribution Monitoring

Public welfare programs, including subsidies, social security payments, and financial aid distributions, are highly susceptible to fraudulent claims and identity manipulation. The proposed framework enables continuous monitoring of beneficiary transactions to ensure that funds are allocated to legitimate recipients.

Through behavioral analytics and anomaly detection, the system can identify duplicate identities, unusual withdrawal patterns, and suspicious eligibility claims. This enhances fairness and accountability in welfare distribution systems and ensures that public resources are utilized effectively for intended beneficiaries.

8.4. National-Level Digital Payment Security Systems

With the rapid expansion of digital payment ecosystems, ensuring secure and fraud-free transactions has become a national priority. The proposed framework can be integrated into national payment gateways and digital financial infrastructures to provide real-time fraud detection and prevention.

By analyzing transaction flows across multiple financial institutions, the system can detect coordinated fraud attempts, money laundering activities, and unauthorized transactions on a national scale. This strengthens the overall security of digital payment ecosystems and enhances public confidence in digital financial services.

9. Future Research Directions

Although the proposed framework provides a comprehensive approach to real-time fraud detection in public financial systems, several promising directions can further enhance its effectiveness, scalability, and transparency. Future research can extend this work in the following areas:

9.1. Blockchain-Based Fraud Prevention Systems

Blockchain technology offers strong potential for enhancing transparency, immutability, and traceability in financial transactions. Future research can explore the integration of blockchain with AI-driven fraud detection systems to create tamper-proof financial records and improve auditability in public financial systems. Chowdhury (2024c) highlights the convergence of artificial intelligence, machine learning, and blockchain in modern financial ecosystems, emphasizing

their combined role in strengthening financial security and operational transparency. Such integration can significantly reduce data manipulation risks and enhance trust in government financial operations.

9.2. Federated Learning for Cross-Agency Fraud Detection

Public financial systems often operate across multiple agencies with fragmented and sensitive datasets. Federated learning presents a promising solution by enabling collaborative model training without sharing raw data. This approach allows different government institutions to jointly detect fraud patterns while preserving data privacy and compliance with regulatory constraints. The integration of distributed AI models aligns with the broader trend of cloud-based scalable analytics systems discussed by Chowdhury (2025a), where decentralized architecture enhances computational efficiency and security in large-scale financial environments.

9.3. Explainable AI (XAI) for Transparency in Government Decisions

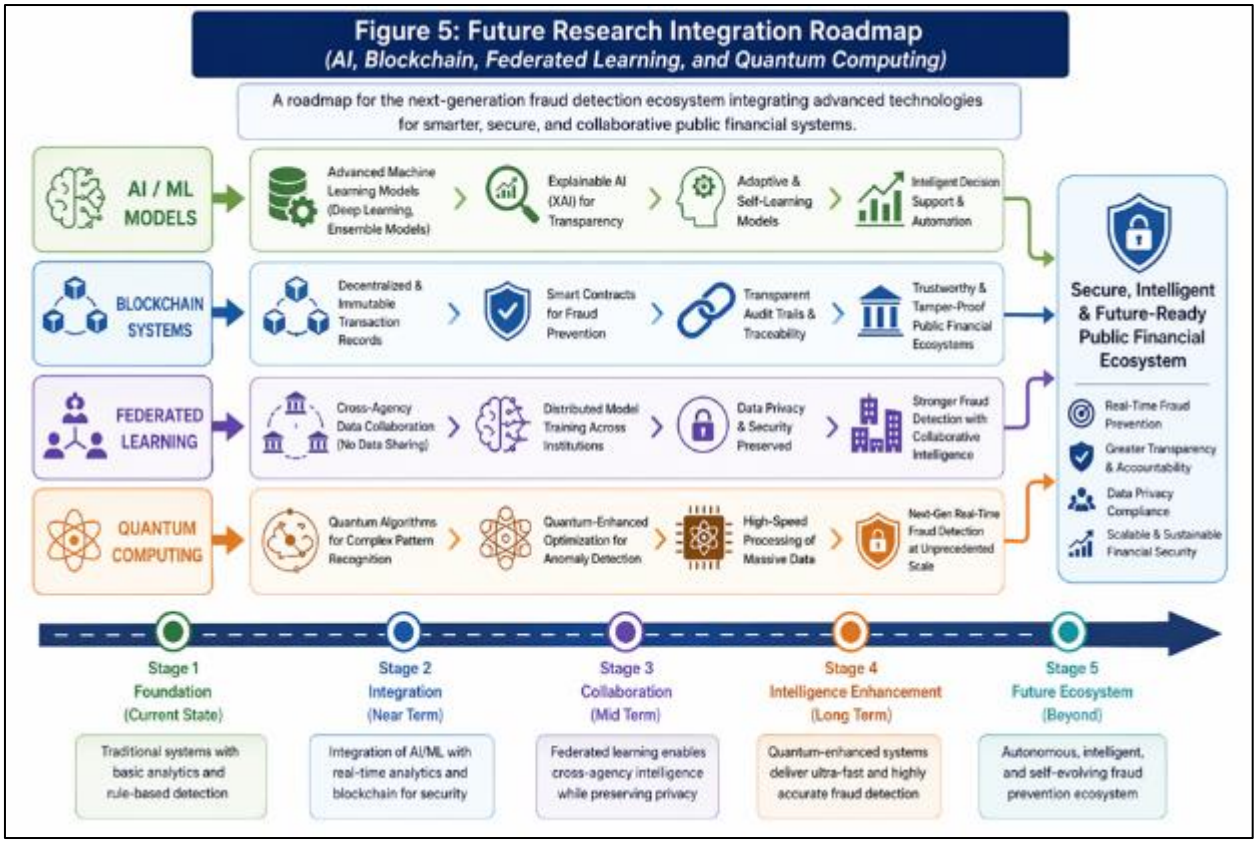
As AI systems become more widely used in financial decision-making, the need for transparency and interpretability becomes critical, especially in public sector applications. Explainable AI (XAI) can help financial authorities understand how fraud detection models arrive at specific decisions, thereby increasing trust and accountability. Future research should focus on developing interpretable machine learning models that provide clear explanations for fraud classifications while maintaining high predictive accuracy. This is particularly important for regulatory compliance and ethical governance in public financial systems.

9.4. Quantum-Enhanced Fraud Detection Models

Quantum computing represents an emerging frontier with the potential to revolutionize fraud detection by significantly accelerating complex data processing and optimization tasks. Quantum-enhanced models may enable faster pattern recognition in large-scale financial datasets and improve the efficiency of anomaly detection algorithms. Although still in the early stages of development, this direction holds promise for future high-performance fraud detection systems capable of handling extremely large and complex financial networks.

9.5. Future Research Integration

Overall, future research in fraud detection for public financial systems is expected to move toward more secure, decentralized, transparent, and computationally advanced frameworks. The integration of blockchain, federated learning, explainable AI, and quantum computing will collectively contribute to building next-generation intelligent financial security systems capable of addressing increasingly sophisticated fraud threats.



(This figure presents a forward-looking roadmap illustrating the progressive integration of advanced computational paradigms in next-generation intelligent systems. It highlights the convergence of Artificial Intelligence and Machine Learning models with Blockchain-based distributed systems and Federated Learning frameworks, enabling secure, decentralized, and privacy-preserving analytics. The roadmap further extends toward the anticipated role of Quantum Computing as an enabling technology for exponential computational scalability and optimization. Overall, the layered structure demonstrates the staged evolution from current AI-driven architectures to a fully integrated, hybrid ecosystem supporting trustworthy, scalable, and distributed intelligence for future research and applications.)

Figure 5 Future Research Integration Roadmap (AI, Blockchain, Federated Learning and Quantum Computing)

10. Conclusion

This study proposed a comprehensive big data analytics framework for real-time fraud detection in public financial systems, integrating stream processing, machine learning, and distributing computing technologies. The framework was designed to overcome the limitations of traditional fraud detection approaches by enabling continuous monitoring, rapid anomaly detection, and adaptive learning from high-volume financial transactions. Through the combination of supervised, unsupervised, and hybrid AI models, the system provides a robust mechanism for identifying both known and emerging fraud patterns in dynamic public financial environments.

The contribution of this research to public financial security is significant, as it offers a scalable and intelligent solution for mitigating fraud risks across critical government financial operations. By enhancing detection accuracy, reducing false positives, and enabling real-time decision-making, the proposed framework strengthens the integrity and transparency of public financial systems. It also supports improved governance by enabling authorities to respond quickly to suspicious activities and prevent financial losses at an early stage.

Furthermore, this study highlights the growing importance of real-time analytics in modern governance systems. As governments increasingly adopt digital financial infrastructures, the need for immediate data processing and intelligent decision-making becomes essential. Real-time analytics not only improve operational efficiency but also enhances accountability, trust, and security in public financial management. Overall, the integration of big data analytics and AI-driven technologies represents a transformative step toward building more secure, responsive, and resilient public financial ecosystems.

References

[1] Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. Journal of Network and Computer Applications, 68, 90–113. https://doi.org/10.1016/j.jnca.2016.04.007

[2] Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. International Conference on Computing Networking and Informatics, 1–9.

[3] Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), 602–613. https://doi.org/10.1016/j.dss.2010.08.008

[4] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58. https://doi.org/10.1145/1541880.1541882

[5] Chowdhury, R. H. (2024a). Advancing fraud detection through deep learning: A comprehensive review. World Journal of Advanced Engineering Technology and Sciences, 12(02), 606–613.

[6] Chowdhury, R. H. (2024b). Harnessing machine learning in business analytics for enhanced decision-making. World Journal of Advanced Engineering Technology and Sciences, 12(02), 674–683.

[7] Chowdhury, R. H. (2024c). Public debt management with advanced data analytics. AI and Data Science Journal, 1(01), 6–19.

[8] Chowdhury, R. H. (2024d). The evolution of business operations: Unleashing the potential of artificial intelligence, machine learning, and blockchain. World Journal of Advanced Research and Reviews, 22(03), 2135–2147.

[9] Chowdhury, R. H., Masum, A. A., Farazi, M. Z. R., & Jahan, I. (2024). The impact of predictive analytics on financial risk management in businesses. World Journal of Advanced Research and Reviews, 23(03), 1378–1386.

[10] Chowdhury, R. H., Prince, N. U., Abdullah, S. M., & Mim, L. A. (2024). The role of predictive analytics in cybersecurity: Detecting and preventing threats. World Journal of Advanced Research and Reviews, 23(02), 1615–1623.

[11] Chowdhury, R. H. (2025a). Cloud-based data engineering for scalable business analytics solutions: Designing scalable cloud architectures to enhance the efficiency of big data analytics in enterprise settings. Journal of Technological Science & Engineering, 2(1), 21–33.

[12] Chowdhury, R. H. (2025b). Digital engagement and customer experience leadership: Strategies for the modern enterprise. Deep Science Publishing.

[13] Chowdhury, R. H. (2025c). Digital leadership and organizational learning: Technologies for business transformation and operational excellence. Deep Science Publishing.

[14] Chowdhury, R. H. (2025d). Utilizing business analytics to combat financial fraud and enhance economic integrity. International Journal of Science and Research Archive, 14(01), 134–145.

[15] Dahiya, S., & Bhatia, S. (2021). Big data analytics in financial fraud detection: A systematic review. Procedia Computer Science, 182, 215–222. https://doi.org/10.1016/j.procs.2021.03.024

[16] Huang, D., & Li, W. (2019). Real-time fraud detection in financial systems using big data analytics. Future Generation Computer Systems, 95, 65–74. https://doi.org/10.1016/j.future.2018.12.030

[17] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review. Decision Support Systems, 50(3), 559–569. https://doi.org/10.1016/j.dss.2010.08.006

[18] Sharma, S., & Panigrahi, P. K. (2013). A review of financial accounting fraud detection based on data mining techniques. International Journal of Computer Applications, 39(1), 37–47. https://doi.org/10.5120/4841-7034

[19] Thota, S. R., & Kim, J. (2020). Streaming analytics for real-time fraud detection in financial systems. IEEE Access, 8, 121321–121334. https://doi.org/10.1109/ACCESS.2020.3006321

[20] West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. Computers & Security, 57, 47–66. https://doi.org/10.1016/j.cose.2015.11.005

[21] Zhang, Y., Wang, S., & Ji, G. (2019). A comprehensive survey on data mining for fraud detection in financial services. IEEE Access, 7, 125503–125523. https://doi.org/10.1109/ACCESS.2019.2933335