Received: 7 May 2025 | Revised: 22 July 2025 | Accepted: 11 August 2025 | Published online: 1 September 2025

RESEARCH ARTICLE

FinTech and Sustainable Innovation
2025, Vol. 00(00) 1–9
DOI: 10.47852/bonviewFSI52026108

BON VIEW PUBLISHING

Credit Card Fraud Detection Through Deep Learning and Real-Time Data Streams: A Comparison and New Directions

Elias Polytarchos[1,*]

[1]*Department of Management Science & Technology, Athens University of Economics & Business, Greece*

Abstract: Credit card fraud detection has become a critical concern for financial institutions as the volume of digital transactions grows rapidly. This paper presents a comparative study of two advanced methodologies – deep learning and real-time data stream analysis – applied to credit card fraud detection. Deep learning models, particularly Long Short-Term Memory networks, demonstrated high accuracy (up to 92%) in predicting customer behaviors and contributing to the detection of fraudulent transactions. However, they require large amounts of historical data and may not be ideal for real-time detection. In contrast, real-time data stream analysis, powered by innovative techniques like the patented BEReTiC system, provides immediate fraud detection but with lower initial accuracy. This paper explores the trade-offs between these approaches, highlighting the strengths of deep learning in pattern recognition and the adaptability of real-time data mining in dynamic financial environments. We evaluate both techniques on real-world data, measuring accuracy, false positives, and adaptability to novel fraud patterns. Results indicate that while deep learning offers high accuracy, BEReTiC enables faster detection with fewer false alarms and enhanced responsiveness. The findings suggest that a hybrid model integrating both techniques may offer a more effective solution for tackling the complexities of credit card fraud in real time, since it would combine the predictive power of deep models with the agility of real-time analytics, opening new directions in fraud prevention for high-velocity financial environments.

Keywords: clustering algorithms, data streams, pattern recognition, time series analysis, deep learning, fraud detection

1. Introduction

In the modern era of digital transactions, credit card fraud has emerged as a significant challenge for financial institutions and consumers alike. As online payments and e-commerce activities continue to rise, credit card fraud remains a persistent challenge due to the sophistication of evolving fraud tactics and the imbalance inherent in transaction datasets. This growing threat not only causes substantial financial losses but also affects consumer trust and compliance with regulatory obligations, making effective detection systems a critical operational priority. Consequently, there is a need for more advanced, scalable solutions that can analyze data in real time, ensuring quicker detection of fraudulent transactions.

Traditional methods, such as rule-based systems and machine learning models trained on historical data, face limitations in real-time fraud detection, where immediate action is critical. While deep learning techniques like Long Short-Term Memory (LSTM) networks excel in identifying intricate patterns within time series data, they require extensive preprocessing and retraining, making them less practical for dynamic environments. Conversely, real-time clustering methods, such as the BEReTiC system, are designed to detect anomalies instantly by analyzing data streams without prior cleansing.

Despite the emergence of such tools, existing literature lacks a comprehensive, empirical comparison of real-time unsupervised stream-based methods with deep learning-based batch-trained models, particularly in fraud detection scenarios involving highly imbalanced, high-volume financial data. This gap is critical: without such comparative analyses, system designers cannot make informed decisions about deploying timely and effective fraud prevention mechanisms.

This paper presents a comparative evaluation of these methodologies, exploring their performance on a large, real-world dataset of credit card transactions. The study focuses on addressing critical gaps in speed, adaptability, and accuracy to propose a hybrid solution that combines the strengths of both approaches. In particular, it aims to evaluate the effectiveness of the patented BEReTiC real-time data stream analysis methodology, which is capable of operating without prior data cleansing, against high-performing deep learning frameworks trained on historical data. To our knowledge, this is among the first comparative evaluations of deep learning approaches against real-time data stream frameworks for fraud detection. The deep learning methods are used only for comparison; the research does not claim any contribution to these.

The core research question explored in this paper is: Can real-time pattern recognition without any a priori knowledge be employed to detect fraud on data streams of transactions, and how does it fare against batch-trained deep learning?

***Corresponding author:** Elias Polytarchos, Department of Management Science & Technology, Athens University of Economics & Business, Greece. Email: ipoli@aueb.gr

The BEReTiC methodology, the SCoDe2 mechanism, and the CluNN algorithm presented in this paper are the object of the patent PCT/GR2024/000039 / 02-12-2024 (GR 1010876) [1].

2. Background

Typically, fraud detection in credit card transactions and financial risks in general is considered a data mining problem, where clustering algorithms and other unsupervised machine learning methods are employed to analyze datasets and detect suspicious actions [2–6].

An interesting approach, in the context of credit card fraud detection, where the number of fraudulent transactions is significantly lower than non-fraudulent ones, leading to imbalanced datasets, is Generative Adversarial Networks (GANs) [7]. Traditional binary classification models may struggle with imbalanced data, as they tend to bias results toward the majority class; while oversampling the minority class is a common technique to address imbalance, it has its limitations. GANs are deep learning models that are trained to generate synthetic examples of the minority class, creating a more balanced training set [8]. The generated minority class examples are combined with the original dataset, creating an augmented training set. Experiments show that classifiers trained on the augmented set outperform those trained on the original data. Notably, sensitivity, which is crucial in fraud detection, is significantly enhanced. The combined approach of GANs and augmented training sets results in a more effective fraud detection mechanism [9].

In financial applications, akin to numerous real-world scenarios, the data poses challenges that are less prevalent in traditional academic datasets. Notable among these challenges are issues related to size, noise, sparsity, and uncertainty. Moreover, the majority of financial datasets exhibit a pronounced imbalance [10]. Take credit card applications, for instance, where, as noted above, the number of reliable customers significantly outweighs that of problematic customers. Similarly, in fraud detection, the dataset is predominantly composed of normal transactions, with only a sparse representation of fraudulent ones. Consequently, there arises a need for predictive analytics techniques adept at handling the intricacies of unbalanced financial datasets, facilitating the creation of accurate and interpretable financial models [11]. A wealth of research exists regarding the classification of unbalanced datasets. An excellent review, which does not focus on the real-time facet of the issue, can be found at [12]. On the other hand, it presents a comprehensive study on the application of machine learning techniques for real-time fraud detection in financial transactions [13]. It explores both supervised and unsupervised models (e.g., decision trees, Support Vector Machines, and neural networks) to identify patterns and anomalies indicative of fraud. The study highlights the integration of these models into scalable systems capable of handling high-transaction volumes with low latency and improved accuracy. Experimental results demonstrate significant improvements in fraud detection rates and false positive reduction.

Anomaly detection is a cornerstone of financial fraud detection, where malicious behavior often appears as subtle deviations in high-volume, high-dimensional, or sequential transaction data. The Empirical-Cumulative-distribution-based Outlier Detection (ECOD) methodology provides a lightweight, parameter-free, and interpretable approach by identifying rare events through empirical cumulative distributions, facilitating scalable and transparent detection [14]. Score-Guided Networks (SGN) [15] further enhance performance by learning scoring patterns from existing detectors in a teacher-student setup, improving accuracy without domain-specific tuning. Complementing these, rule-based approaches for anomaly detection in sequence data [16] introduce symbolic representations and interpretable rule violations, offering valuable insights into time-dependent financial streams. Together, these methodologies exemplify how unsupervised learning – grounded in efficiency, accuracy, and explainability – can form the foundation of adaptive fraud detection systems, with strong potential for integration into near-real-time analytics pipelines.

Beyond specialized fraud detection efforts, advancements in real-time financial modeling offer valuable conceptual parallels. For example, Zhang et al. [17] introduced a cost-sensitive deep reinforcement learning approach for portfolio optimization, addressing the challenges of non-stationary environments and asymmetric transaction costs – both of which are fundamental to fraud analytics. Similarly, Li et al. [18] present a multimodal LSTM architecture that integrates temporal and textual signals for event-driven prediction in financial markets. While the aim of the paper is stock forecasting, the underlying mechanisms (i.e., sequence modeling, concept drift adaptation, and multimodal fusion) resonate strongly with the objectives of real-time, data-driven fraud detection systems.

In the next sections, we will present the dataset employed, as well as analyze the employed methodology and present the results.

3. Dataset

Two separate datasets that contained historical data for a single year were provided:

IND dataset: 17514242 individual credit/debit card transactions and

SUM dataset: 1207817 summaries for credit/debit card purchases

The data were labeled, complete, and well structured. The IND dataset contained a non-negative time series, where each record, besides the timestamp of the purchase and the customer index, also contained information regarding the specific transaction (merchant type, coarse location, month, season, and price range), as well as demographic information (whether the customer has children or not, gender, education, occupation category, age range, total funds range). The SUM dataset contained the total number of purchases of every customer separated by category of expense (automotive and maritime, gambling, energy, government and taxes, health, home equipment, insurance, education, leisure activities, dining, clothing, services, FMCG, electronics, travel) along with their total amount and demographic information (the demographic information included in the SUM dataset was the same as IND, with the addition of education level, marital status, city, customer type).

4. Methodology

Traditional approaches to fraud detection primarily frame the problem as a data mining challenge, often dealing with unbalanced datasets where fraudulent transactions represent a tiny fraction of the data. While these approaches have been somewhat effective, they come with several inherent limitations when applied to real-world, dynamic environments. Specifically, traditional methods rely on:

Machine learning on unbalanced data: These models must contend with the imbalance of legitimate versus fraudulent transactions, requiring strategies such as oversampling, undersampling, or synthetic data generation. However, these techniques can fall short in adapting to emerging fraud patterns.

Vetted, pre-cleansed datasets: For accuracy, these models often depend on pre-processed, well-curated datasets, a step that not only introduces delays in fraud detection but also limits the adaptability of the models to real-time environments.

Frequent retraining: Machine learning models require regular retraining to recognize new types of fraud, which incurs a significant cost in terms of both time and resources. This retraining cycle can hinder the detection of rapidly evolving fraud tactics in real time.

These limitations make conventional methodologies less suitable for real-time fraud detection, where transactions must be analyzed instantly to prevent fraudulent activities as they occur. In this paper, we propose a real-time clustering approach that directly addresses these shortcomings and represents a novel methodological contribution to the field.

Our key innovation lies in the development of a real-time clustering methodology designed to detect anomalies in credit card transactions as they happen. This approach diverges from traditional methods in several crucial ways:

Dynamic clustering on live data: Instead of relying on historical, pre-cleansed datasets, our system operates on live transaction streams. Transactions are clustered dynamically based on their features, such as transaction amount, location, time, and frequency. By continuously updating these clusters in real time, our model is able to identify transactions that deviate from established patterns and flag them for further investigation.

Immediate detection of novel fraud patterns: Traditional methods can struggle to detect new fraud tactics until they have occurred, and the models have been retrained. In contrast, our real-time clustering approach is inherently adaptive, capable of identifying emerging patterns without needing retraining. Any significant deviation from existing clusters is immediately flagged as suspicious, allowing for faster fraud intervention [19].

Scalability and adaptability: Unlike deep learning and traditional machine learning models, which can be resource-intensive and require significant computational power for retraining, our clustering method is lightweight and highly scalable. This makes it particularly well-suited for environments with large volumes of transactions, such as those handled by financial institutions. Our system can analyze and adapt to patterns in real time, even as the volume of transactions fluctuates.

No need for data cleansing or preprocessing: One of the major bottlenecks in fraud detection is the time spent cleansing and vetting datasets to ensure models are accurate. Our real-time clustering methodology bypasses this requirement, as it works directly on raw transaction data, making it far more efficient in practical deployment. This allows for quicker detection and response to fraud, without the overhead of preparing datasets.

To assess the effectiveness of our real-time clustering method, we also developed a deep learning-based approach as a benchmark for comparison. Deep learning models, known for their ability to identify complex patterns, were applied to the same dataset used for real-time clustering. However, despite their powerful capabilities, the deep learning models required extensive preprocessing and retraining, making them less suitable for real-time fraud detection. The comparison highlights the capability of real-time clustering to deliver acceptable results in environments where immediate detection is crucial but also exemplifies the excellent results that can be achieved through finely tuned deep learning methodologies.

To evaluate fraud detection capabilities, we injected synthetic fraudulent transactions into the IND dataset, which contains individual transaction records. The first step was to define what constitutes a fraudulent transaction. A transaction conducted by someone other than the rightful party is clearly fraudulent; however, detecting such behavior requires the presence of additional indicators. These include deviations from the customer's typical transaction profile, as inferred from attributes such as merchant type, general location, time of year (month and season), price range, and the recorded demographic characteristics of the transaction (children, gender, education, occupation category, age range, and total funds range). We also incorporated aggregated behavioral data from the SUM dataset, specifically the distribution of purchase categories per customer.

Based on this, we performed an initial processing step where we averaged the transactional features of each customer. Using these profiles, we generated 1000 synthetic transactions that aligned with the average behavior of a different customer than the one they were assigned to – thereby simulating fraudulent behavior based on behavioral inconsistency.

The main contribution of this paper is the introduction of a real-time fraud detection system based on dynamic clustering. This method offers several advantages over traditional and deep learning-based approaches, including:

Speed: Instantaneous detection of fraud without the need for data cleansing or retraining.

Adaptability: The ability to identify novel fraud patterns in real time, addressing the rapidly evolving nature of fraud tactics.

Scalability: Efficient operation in high-transaction environments with minimal computational overhead.

By addressing the limitations of existing fraud detection methods, our real-time clustering approach presents a novel solution that is both practical and effective for high-velocity transaction systems.

To address the challenges of fraud detection in high-velocity financial environments, we implemented and evaluated two distinct approaches: a deep learning-based pipeline and a real-time data stream analysis system. This section outlines the methodology followed in this comparative study, broken down into two main branches:

Deep learning pipeline: This approach involved preprocessing a large historical dataset, training multiple supervised models (LSTM and MLP) to classify key demographic and behavioral labels of the customer and computing the degree to which transactions deviate from the expected profile. The deviation was quantified using a composite metric (SST), which enabled the identification of potentially fraudulent transactions based on misclassification consensus among the ensembles.

Real-time clustering and classification: In parallel, we deployed a real-time system (BEReTiC) capable of processing streaming data without prior cleansing. This system dynamically clusters transactions and customer profiles using a semi-supervised approach, detects behavioral deviations in real time, and flags anomalies through concept drift and similarity analysis (Gower metric).

In both approaches, synthetic fraudulent transactions were generated and injected into the dataset to assess detection capability. The deep learning models operated in a batch setting, while the real-time pipeline processed data incrementally, adapting to emerging patterns. The subsequent sections describe these methodologies in detail, along with the experimental design and evaluation criteria.

4.1. Deep learning

Inspired by the paradigm of ensemble classifiers and the fact that each credit card transaction is performed between two known bank entities (customer and business), for whom the bank has ample information, we used the approaches described below.

Our initial approach focused on employing state-of-the-art machine learning methodologies to predict the total funds category of the customer that performed the transactions, based on the IND dataset. The total funds category was selected due to its capability to be used in fraud detection or targeted advertisements. The higher possibility of a customer who will consistently be classified into a higher total funds category can indicate that they (a) are implicated

in fraud; (b) can exceed their credit limit, increasing the probability of having to pay interest; and/or (c) are interested in specific categories of advertisements. As the allegation of fraud cannot be based on the misclassification of a model, however accurate this might be, the case of a fraudulent transaction should be based on firmer grounds. In order to do this, our approach was augmented by expanding the usual ensemble classifier scheme. Instead of using the ensemble to classify an object into a class, we used the ensemble to classify an object into multiple classes and then verify that the predicted classes were correctly selected for this object.

Concisely, the models learn the patterns of the various transactions being performed by each customer, and then, by calculating how probable it is that the customer would have normally performed the transaction, they are able to collectively flag a transaction as suspicious. Below, we define the method that accomplishes that:

Each customer can be classified using a number of labels: whether the customer has children or not, gender, occupation category, age range, total funds range, education level, marital status, city, and customer type.

Each transaction can also be classified using the following labels: merchant type, coarse location, and price range.

Each transaction is being performed by a debit or credit card that belongs to a customer: As such, every transaction inherits the labels of the customer.

A number of classifiers are trained using the dataset: In order to be able to classify a transaction into each of these labels.

We defined the Scale of Suspicious Transaction (SST): The percentage of the classifiers that did not correctly predict the relevant labels of the customer that performed a transaction and the Scale of Legitimate Transaction (SLT) as the complement of the SST.

We defined the Confidence of the SST (CSST): The product of the accuracy that had been achieved during the training of the classifiers that comprised the SST and the Confidence of the SLT (CSLT) as the relevant value for the SLT.

Consequently: If the SST is larger than 0.5 (i.e., the majority of the methods in the ensemble misclassified the transaction), and the CSST is larger than the CSLT, then the transaction is considered as a possible fraud with a confidence factor of CSST, and thus, a large amount of transactions flagged as possible fraud within a short period of time signifies a high probability of fraud.

The method above aims to determine, through this "fuzzy" classification of a transaction to labels that match the labels assigned to the customer, the probability that a transaction could have been performed by this customer under standard conditions. A mounting number of transactions of high SST and CSST values can indicate fraudulent movements [20]. We employed two deep learning methods for the models using the TensorFlow infrastructure. Specifically:

We created a model that used an LSTM Recurrent Neural Network, since the dataset was a time series of transactions and LSTMs have been shown to work well with time series [18, 21–24].

We also performed experiments using simple deep sequential multilayer perceptron models (MLP) [25].

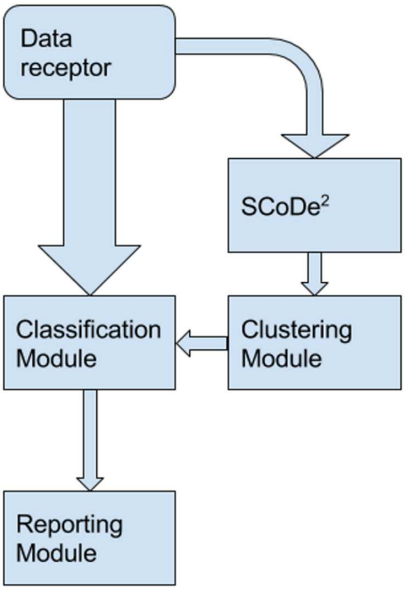
After training the models, we employed them on a dataset that included only legitimate transactions (according to the bank) and afterward on the same dataset enriched with some that would have to be classified as possibly fraudulent.

4.2. Real time

Credit card transactions can be considered as a data stream – a stochastic time series – that can contain patterns. By being able to detect and analyze patterns in real time, a multitude of new possibilities could emerge, for example: (a) better and more efficient protection of a bank's customers from theft, (b) real-time detection of fraudulent transactions, (c) ability to support gamified challenges for more playful customers (e.g., by automatically notifying customers and providing rewards for transactions performed according to a newly detected trend) [26], (d) automatic deduction of appropriate advertisements for the customers, or (e) fast response to unprecedented events.

To tackle real-time pattern recognition, we employed our work in [1], where we defined a system that can analyze the data stream in real time using a semi-supervised machine learning methodology (Best Effort Real-Time Clustering and Classification adapter – BEReTiC). This contains the modules described below. Their relations are depicted in Figure 1 [1].

Figure 1
Relations of the BEReTiC modules



Data receptor: This module undertakes the reception and initial parsing of the data, discarding any items that cannot be properly parsed. It does not perform any filtering or preprocessing of the data; it maps the data to a data structure.

Sample collector and deviation detector (SCoDe^2): The SCoDe^2 monitors the stream, stores a representative sample of the received detections, measures the statistical properties (standard deviation), and calculates and keeps track of the mode of distances between the entities.

Clustering module: The clustering module detects clusters on the sample collected by SCoDe^2, using the calculated mode of distances as an additional input. The clustering module processes the sample whenever input is available from the sample collector, initiating a new clustering round. A concept drift gets detected whenever the clusters differ from the clusters of the previous round.

Classification module: Every data point, whether it was used in the context of the previous modules or not, is also classified by the classification module in one of the detected clusters.

Reporting module: As the designed system is a web service, it does not contain a user interface; this module has been designed

in order to provide information, such as the statistical properties of the stream, the clusters detected, and the number of concept drifts.

By converting the IND dataset into a data stream and feeding it to the BEReTiC, we were able to detect the patterns of the transactions being performed in real time. The clustering module and the classification module were configured to use the CluNN algorithm [1] and KNN [27], respectively.

The data structure employed by the system included, for each customer:

Their demographic information

A fingerprint of their transactions (the customer's transaction profile)

The mode of the amount of the total transactions

The percentage of transactions per merchant type

The mode of the amount of transactions per merchant type

The percentage of transactions per location

The mode of the number of transactions per location

The percentage of transactions per month

The mode of the number of transactions per month

The percentage of transactions per merchant type per location

The mode of the amount of transactions per merchant type per location

The percentage of transactions per merchant type per month

The mode of the amount of transactions per merchant type per month

The percentage of transactions per location per month

The mode of the number of transactions per location per month

The similarity metric used, which compared instances of the data structure described above, was the Gower similarity [28]. This was selected because it allows the combination of categorical (demographic information) and numerical (the fingerprint of the transactions) data.

Using the aforementioned metric, we were able to compare transactions, customers' fingerprints, and clusters with other transactions, customers' fingerprints, or clusters.

The comparisons above were used to create clusters of transactions (i.e., collections of similar transactions) and fingerprints of customers (i.e., collections of transactions performed by the same customer). As such, by employing the BEReTiC, we were able to dynamically create and update the aforementioned clusters and fingerprints.

Regarding the example insights mentioned at the beginning of the section:

Protection of bank customers from theft and real-time detection of fraudulent transactions. A big dissimilarity of a transaction from the respective customer's fingerprint can indicate a fraudulent transaction.

Ability to support gamified challenges for more playful customers (e.g., by automatically notifying customers and providing rewards for transactions performed according to a newly detected trend). New trends can be considered as new clusters of transactions, and challenges could be proposals to customers to perform transactions that are classified in this cluster.

Automatic deduction of appropriate advertisements for the customers. Using the most common recent transactions of the customers, advertisements for products or services that exist in the same cluster can be proposed.

Fast response to unprecedented events. An example of this is the gamified challenges mentioned above, and these can be enabled through new clusters of transactions and respective actions upon their detection.

5. Results

5.1. Deep learning

Our initial approach, which was designed to predict the field total funds range, achieved a 92% accuracy.

The results of the extended classifier scheme were promising and allowed us to predict each one of the labels with accuracies of up to 92%. Specifically, when using LSTM models, the accuracies were:

- total funds range: 92%
- age range: 86%
- children: 82%
- occupation category: 79%
- customer type: 79%
- education level: 75%
- gender: 73%
- marital status: 72%
- city: 59%

When using MLP models, the accuracies were:

- total funds range: 78%
- age range: 81%
- children: 64%
- occupation category: 71%
- customer type: 64%
- education level: 67%
- gender: 72%
- marital status: 64%
- city: 47%

After injecting the fraudulent transactions, both in the training dataset and the testing dataset, 788 out of the 1000 injected fraudulent transactions were able to be detected, while 1340 false alarms were also classified as possibly fraudulent, that is, transactions that were contained in the original IND dataset, a less than 0.007% misclassification rate (assuming that the provided dataset did not contain additional illegal transactions).

5.2. Real time

When using real-time analysis, the overall accuracies, after feeding the entire dataset, were:

- total funds range: 66%
- age range: 53%
- children: 57%
- occupation category: 62%
- customer type: 51%
- education level: 63%
- gender: 44%
- marital status: 48%
- city: 51%

After injecting the fraudulent transactions, 619 out of the 1000 injected fraudulent transactions were able to be detected, while 574 false alarms were also classified as possibly fraudulent, achieving a 0.003% misclassification rate.

The accuracy of the predictions performed by the real-time approach was substantially lower than the relevant predictions of the deep learning methods. This can be attributed to the fact that the real-time approach was not aware of all the transactions that the customer had historically performed, at least while the system had not

calculated a representative fingerprint of the customer. We define the representative sample of the customer as the one that does not differ from the one obtained when all their transactions have been processed. However, even after having calculated a representative fingerprint of the customer, the results could not reach the accuracy of the deep learning methods.

5.3. New directions

Despite the real-time approach's modest initial performance, it opens several promising avenues for advancing fraud detection systems. Building on the comparative insights of deep learning versus streaming methods presented above, future work can focus on harnessing real-time capabilities to enhance fraud mitigation, improve financial service responsiveness, and increase adaptability to emerging threats. Key directions include technical innovations to blend and extend current models, addressing regulatory and ethical considerations in live decision-making and pioneering application-level features that engage and protect users. Below, we outline these new directions and their potential benefits, challenges, and opportunities for exploration.

5.3.1. Technical opportunities

Hybrid Model Integration: A clear path forward is combining the high accuracy of deep learning with the immediacy of streaming analytics. For example, a hybrid system might use an offline-trained LSTM to provide a strong baseline for fraud scoring, while an online clustering component (e.g., the BEReTiC system) adapts to live data drift. This combination can mitigate the weaknesses of each approach, as deep learning models supply well-learned patterns, while real-time modules adjust to new fraud tactics on the fly. Research into transfer learning or meta-learning for real-time models (to pre-seed streaming detectors with knowledge from deep models) is one promising area to address cold-start accuracy drops. Overall, hybrid architectures could drastically improve fraud mitigation speed (catching fraud in-flight) without sacrificing the pattern-recognition prowess of deep networks.

Adaptive and Self-Learning Systems: A strength of real-time analytics is adaptability, that is, the ability to update the model as new data arrives, thus handling concept drifts in fraud patterns [29]. Future systems could leverage online learning algorithms and reinforcement learning to continuously adjust fraud decision policies. For instance, the BEReTiC can detect trends as new transactions stream in, thereby reflecting current customer behavior. The BEReTiC approach, being a semi-supervised clustering and classification adapter, is potentially more interpretable (clusters of behavior or "fingerprints" can be visualized) than a complex deep neural net, which is a bonus for explaining fraud decisions. Future work can build on this by integrating explainable AI techniques (like rule extraction or prototype examples for clusters) directly into the streaming pipeline.

5.3.2. Application-level innovations

Despite the fact that fraud detection through real-time data stream processing cannot be relied on, as BEReTiC's clustering module continuously groups streaming data points into evolving clusters, it is effectively performing dynamic segmentation of financial behaviors (e.g., spending patterns). If a new pattern of behavior appears (e.g., a surge in a previously rare transaction type or a sudden shift in spending habits), the algorithm will form a new cluster to represent it. The system inherently flags such events as concept evolutions – essentially alerting that a new behavioral pattern or trend has emerged. This mechanism allows BEReTiC to "sense" trends in real time: a cluster that did not exist an hour ago but consistently appears now is a strong signal of an emerging trend. Notably, BEReTiC is designed to work on raw data streams (e.g., transaction logs) without offline cleansing, enabling instant detection of anomalies or novelties. In contrast, traditional methods might only catch these shifts in a later batch analysis (if at all), by which time the trend could be well underway. By maintaining up-to-date clusters, the system effectively produces a constantly revised segmentation of customers or transactions, mirroring the latest patterns in the data stream.

5.4. Summary and discussion of metrics

To improve clarity, Tables 1, 2, and 3 present a consolidated summary of the results for each methodology. While no publicly available implementation of existing systems was compatible with the provided proprietary dataset for direct benchmarking, the goal of this study was to explore the relative merits of two fundamentally different fraud detection paradigms, that is, batch-trained deep learning versus real-time clustering on streaming data. The presented results are therefore meant to offer a baseline for future comparative work and hybrid designs, rather than to claim direct superiority over existing solutions.

Existing literature suggests that deep learning systems, especially when combined with techniques like oversampling or GAN-based augmentation, can achieve high sensitivity in imbalanced fraud detection contexts [7, 9, 30]. However, they often require significant preprocessing and are unsuitable for real-time environments. On the other hand, real-time, unsupervised clustering methods typically prioritize responsiveness and adaptability over accuracy. Our results confirm this trend and highlight a key trade-off: while deep learning achieves higher accuracy in classification, the BEReTiC-based real-time system demonstrates faster fraud response and reduced false alarm rates when working with raw data streams.

The implications of this study resonate beyond the immediate scope of fraud detection and intersect with broader advances in real-time financial analytics. For instance, Manoharan et al. [13] directly address the challenges of machine learning-based fraud detection in high-throughput transactional systems, highlighting the practical feasibility of real-time deployments. Similarly, ECOD [27] introduces a parameter-free and interpretable outlier detection technique based on empirical cumulative distributions, offering scalability and transparency – two properties vital in financial domains. The SGN framework [15] enhances detector accuracy by learning scoring distributions and could complement fraud detection by refining anomaly decision boundaries. Furthermore, the anomaly rule detection approach by Gan et al. [16] provides a mechanism for identifying sequential irregularities in symbolic form, aligning with time-evolving fraud scenarios while ensuring human interpretability. Beyond fraud-specific applications, adjacent works such as [17] apply real-time deep reinforcement learning and multimodal LSTM architectures, respectively, for dynamic portfolio selection and stock prediction. These methods share key objectives with fraud detection: timely decision-making, cost-sensitive optimization, and adaptation to non-stationary data streams. Taken together, these contributions underscore the growing convergence of anomaly detection, real-time learning, and financial decision-making and suggest a fertile landscape for future cross-domain innovations.

The practical deployment of the proposed method in large-scale financial systems also merits consideration. In such environments, transaction volumes can reach thousands per second, necessitating solutions with low inference latency and high throughput.

Table 1
Classification accuracy by label

Label	Deep learning (LSTM)	Deep learning (MLP)	Real-time clustering
Total funds range	92%	78%	66%
Age range	86%	81%	53%
Children	82%	64%	57%
Occupation category	79%	71%	62%
Customer type	79%	64%	51%
Education level	75%	67%	63%
Gender	73%	72%	44%
Marital status	72%	64%	48%
City	59%	47%	51%

Table 2
Fraud detection performance

Metric	Deep learning	Real-time clustering
Injected fraudulent transactions	1000	1000
Detected fraudulent transactions	788	619
False positives	1340	574
Misclassification rate	0.007%	0.003%

Table 3
Methodology trade-offs

Feature	Deep learning	Real-time clustering
Accuracy	High	Moderate
Latency	High (batch processing)	Low (real-time capable)
Preprocessing required	Yes	No
Adaptability to new patterns	Limited (requires retraining)	High (adaptive clustering)
Resource intensity	High	Low to moderate

The modular structure of our real-time framework enables deployment alongside existing stream processing infrastructures (e.g., Apache Flink or Spark Streaming [31]), allowing scalable ingestion and scoring of transactions in near real time. Furthermore, because the clustering and classification components are online and adaptive, they can continuously learn from new data without costly retraining cycles, which is an important consideration in environments affected by concept drift. Overall, the proposed system aligns well with the architectural constraints of modern financial institutions, supporting timely fraud detection without compromising performance or interpretability.

6. Conclusion

This paper explored two advanced methodologies for detecting credit card fraud: deep learning neural networks and real-time data stream analysis using patented techniques. Both methods were applied to a large dataset of credit card transactions, and the results provide insights into the effectiveness of each approach in real-world fraud detection.

The deep learning models, particularly the LSTM networks, demonstrated high accuracy across various classifications, with the most notable results being the prediction of customer "total funds range" at 92% accuracy. This confirms the strength of deep learning in analyzing time series data and identifying suspicious patterns in large, structured datasets. The ability to classify transactions based on customer behavior profiles allowed deep learning models to detect injected fraudulent transactions with a notable success rate, achieving detection for 788 out of 1000 fraudulent cases. However, it should be noted that this approach also yielded a certain number of false positives, which reflects the challenges of accurately identifying fraud in highly imbalanced datasets.

In contrast, the real-time analysis approach, while offering advantages in processing streaming data and detecting fraud as transactions occurred, displayed lower predictive accuracy compared to deep learning. This was primarily due to the cold-start issue, where the system lacked sufficient historical data to make accurate predictions early in the transaction lifecycle. Despite this, the ability to detect fraud in real time and adjust to evolving transaction patterns highlights the practicality of this method in dynamic environments, such as online credit card transactions.

The comparison between these two approaches reveals important trade-offs. Deep learning excels in high-accuracy classification when sufficient historical data is available, but it may not be suitable for real-time applications where immediate fraud detection is critical. On the other hand, real-time clustering techniques are highly adaptable and allow for rapid identification of suspicious transactions, albeit at the cost of lower initial accuracy. Future research might focus on integrating both methodologies to combine the high accuracy of deep learning with the real-time adaptability of clustering, potentially leading to an optimal solution for credit card fraud detection.

The hybrid nature of the proposed future work of integration (i.e., the juxtaposition of deep learning-based detection with real-time data stream models) offers a valuable synthesis of predictive power and operational feasibility. While deep neural networks excel at capturing complex patterns in transactional behavior, their computational cost and need for retraining limit their applicability in real-time, high-volume environments. Conversely, the real-time stream-based framework emphasizes adaptability, low latency, and continuous learning from unlabeled data. By leveraging the strengths of both paradigms, this hybrid strategy enables comprehensive evaluation under practical constraints: deep learning models serve as high-accuracy offline benchmarks, while real-time stream learners are suited for production deployment. This dual approach ensures that performance is not pursued at the expense of responsiveness or scalability, and it opens the door to layered systems where batch-trained models periodically inform or calibrate online models in a continual learning loop.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by the author.

Conflicts of Interest

The author declares that he has no conflicts of interest to this work.

Data Availability Statement

Data are not publicly available as they were provided to the author under a non-disclosure agreement for the purposes of this study.

Author Contribution Statement

Elias Polytarchos: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Polytarchos, E., Bardaki, C., & Pramatari, K. (2024). *Error prediction and predictive memory maintenance through real time machine learning methods*. Retrieved from: https://worldwide.espacenet.com/patent/search?q=pn%3DGR1010876B

[2] Kou, G., Peng, Y., & Wang, G. (2014). Evaluation of clustering algorithms for financial risk analysis using MCDM methods. *Information Sciences*, *275*, 1–12. https://doi.org/10.1016/j.ins.2014.02.137

[3] Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv Preprint:1009.6119*.

[4] Williams, G. J., & Huang, Z. (1997). Mining the knowledge mine: The hot spots methodology for mining large real world databases. *Australian Joint Conference on Artificial Intelligence, 1342*, 340–348. https://doi.org/10.1007/3-540-63797-4_87

[5] Adhikari, P., Hamal, P., & Jnr, F. B. (2024). Artificial Intelligence in fraud detection: Revolutionizing financial security. *International Journal of Science and Research Archive*, *13*(01), 1457–1472. https://doi.org/10.30574/ijsra.2024.13.1.1860

[6] Guggilam, S., Chandola, V., & Patra, A. (2022). Tracking clusters and anomalies in evolving data streams. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *15*(2), 156–178. https://doi.org/10.1002/sam.11552

[7] Mehri, H., Hawkin, J., Nickerson, K. L., Bihlo, A., & Shoeleh, F. (2024). BankGAN: A generative model for synthetic financial transactions. In *The 37th Canadian Conference on Artificial Intelligence*, 1–12.

[8] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ..., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144. https://doi.org/10.1145/3422622

[9] Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, *479*, 448–455. https://doi.org/10.1016/j.ins.2017.12.030

[10] Sundarkumar, G. G., & Ravi, V. (2015). A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance. *Engineering Applications of Artificial Intelligence*, *37*, 368–377. https://doi.org/10.1016/j.engappai.2014.09.019

[11] Antonelli, M., Bernardo, D., Hagras, H., & Marcelloni, F. (2016). Multiobjective evolutionary optimization of type-2 fuzzy rule-based systems for financial data classification. *IEEE Transactions on Fuzzy Systems*, *25*(2), 249–264. https://doi.org/10.1109/TFUZZ.2016.2578341

[12] Pei, W., Xue, B., Zhang, M., Shang, L., Yao, X., & Zhang, Q. (2023). A survey on unbalanced classification: How can evolutionary computation help? *IEEE Transactions on Evolutionary Computation*, *28*(2), 353–373. https://doi.org/10.1109/TEVC.2023.3257230

[13] Manoharan, G., Dharmaraj, A., Sheela, S. C., Naidu, K., Chavva, M., & Chaudhary, J. K. (2024). Machine learning-based real-time fraud detection in financial transactions. In *2024 International Conference on Advances in Computing, Communication and Applied Informatics*, 1–6. https://doi.org/10.1109/ACCAI61061.2024.10602350

[14] Li, Z., Zhao, Y., Hu, X., Botta, N., Ionescu, C., & Chen, G. H. (2022). Ecod: Unsupervised outlier detection using empirical cumulative distribution functions. *IEEE Transactions on Knowledge and Data Engineering*, *35*(12), 12181–12193. https://doi.org/10.1109/TKDE.2022.3159580

[15] Huang, Z., Zhang, B., Hu, G., Li, L., Xu, Y., & Jin, Y. (2023). Enhancing unsupervised anomaly detection with score-guided network. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(10), 14754–14769. https://doi.org/10.1109/TNNLS.2023.3281501

[16] Gan, W., Chen, L., Wan, S., Chen, J., & Chen, C. M. (2021). Anomaly rule detection in sequence data. *IEEE Transactions on Knowledge and Data Engineering*, *35*(12), 12095–12108. https://doi.org/10.1109/TKDE.2021.3139086

[17] Zhang, Y., Zhao, P., Wu, Q., Li, B., Huang, J., & Tan, M. (2020). Cost-sensitive portfolio selection via deep reinforcement learning. *IEEE Transactions on Knowledge and Data Engineering*, *34*(1), 236–248. https://doi.org/10.1109/TKDE.2020.2979700

[18] Li, Q., Tan, J., Wang, J., & Chen, H. (2020). A multimodal event-driven LSTM model for stock prediction using online news. *IEEE Transactions on Knowledge and Data Engineering*, *33*(10), 3323–3337. https://doi.org/10.1109/TKDE.2020.2968894

[19] Kim, H., Kim, S., Min, S., & Lee, B. (2023). Contrastive time-series anomaly detection. *IEEE Transactions on Knowledge and*

Data Engineering, *36*(10), 5053–5065. https://doi.org/10.1109/TKDE.2023.3335317

[20] Large, J., Lines, J., & Bagnall, A. (2019). A probabilistic classifier ensemble weighting scheme based on cross-validated accuracy estimates. *Data Mining and Knowledge Discovery*, *33*(6), 1674–1709. https://doi.org/10.1007/s10618-019-00638-y

[21] Bandara, K., Bergmeir, C., & Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert Systems with Applications*, *140*, 112896. https://doi.org/10.1016/j.eswa.2019.112896

[22] Connor, J. T., Martin, R. D., & Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks*, *5*(2), 240–254. https://doi.org/10.1109/72.279188

[23] Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, *37*(1), 388–427. https://doi.org/10.1016/j.ijforecast.2020.06.008

[24] Liu, Y., Gong, C., Yang, L., & Chen, Y. (2020). DSTP-RNN: A dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction. *Expert Systems with Applications*, *143*, 113082. https://doi.org/10.1016/j.eswa.2019.113082

[25] Zhao, Q., Wang, F., Wang, W., Zhang, T., Wu, H., & Ning, W. (2025). Research on intrusion detection model based on improved MLP algorithm. *Scientific Reports*, *15*(1), 5159. https://doi.org/10.1038/s41598-025-89798-0

[26] Chauhan, S., Akhtar, A., & Gupta, A. (2021). Gamification in banking: A review, synthesis and setting research agenda. *Young Consumers: Insight and Ideas for Responsible Marketers*, *22*(3), 456–479. https://doi.org/10.1108/YC-10-2020-1229

[27] Fix, E., & Hodges, J. L. (1989). Discriminatory analysis, nonparametric discrimination: Consistency properties. *International Statistical Review*, *57*(3), 238–247. https://doi.org/10.2307/1403797

[28] Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, *27*(4), 857–871. https://doi.org/10.2307/2528823

[29] Pelosi, D., Cacciagrano, D., & Piangerelli, M. (2025). Explainability and interpretability in concept and data drift: A systematic literature review. *Algorithms*, *18*(7). https://doi.org/10.3390/a18070443

[30] You, D., Xiao, J., Wang, Y., Yan, H., Wu, D., Chen, Z., . . . , & Wu, X. (2023). Online learning from incomplete and imbalanced data streams. *IEEE Transactions on Knowledge and Data Engineering*, *35*(10), 10650–10665. https://doi.org/10.1109/TKDE.2023.3250472

[31] Carbone, P., Katsifodimos, A., Ewen, S., Markl, V., Haridi, S., & Tzoumas, K. (2015). Apache flink: Stream and batch processing in a single engine. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, *36*(4), 28–38.

How to Cite: Polytarchos, E. (2025). Credit Card Fraud Detection Through Deep Learning and Real-Time Data Streams: A Comparison and New Directions. *FinTech and Sustainable Innovation*. https://doi.org/10.47852/bonviewFSI52026108