Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2021, 8(1):141-150

Research Article ISSN: 2394 - 658X

Multi-Modal GAN for Combining Machine and Financial Stream

Nirup Kumar Reddy Pothireddy

Independent Researcher

ABSTRACT

With the adoption of cyber-physical systems and more automated information structures comes an exponential rise in both industrial and financial data both for utilization and savings. Like in telecommunication, energy metering, smart manufacturing, and SaaS billing—financial menace, such as overbilling, underreporting, and misuse of operational assets—is often buried in way-out-of-array, time-series high-quality data streams. These anomalies typically appear when, upon the unholy matrimony between machine activity logs with intersecting session data—be it CPU usage, runtime metrics, or system logs—there lies a ripper across matched financial transaction records. Many a time a trained eye will observe anomalies present in machine activity/call quarantine with transaction outputs at times when the model has no such provision to consider operational triggers against transactional effects.

To overcome this hurdle, this paper presents a novel multimodal GAN-based model to model and detect financial anomalies originating from distant but correlated machine and financial streams. The model, created from there, mainly makes it possible to fuse two different types of input modality: (1) sequences of machine activity and (2) billing records or transaction logs in time alignment. Our model's essence is the aforementioned introduction: this two-stream encoder element to learn latent space representations of each individual modality and then a fusion module through the shared layers and LSTM as a time-based representation to put together the representations. The adversarial play in the generative component assesses the reconstruction of machine-financial sequences and duplication in discriminating abnormal techniques at the highest-level using differences about domain discrepancies.

Interestingly, this multimodal GAN model is very different from conventional anomaly detectors that regard any anomaly issuing from statistical irregularities or reconstruction errors, while our proposed mechanism gathers the sweet spot of radical GAN contradiction for modeling fine-grained anomalies in the co-distribution space of financial and machine behavior. The scheme excels in subtle and/or gradual anomalies across various instances, just like: overbilling repeated without corresponding increase in machine sweat; underreporting usage in pay-as-you-go cloud services. And by injecting synthetic data creation in GAN training, the system enhances the generalizability of the model while counteracting issues related to data imbalance in limited samples of labeled anomalies.

The framework comprising the combination of real-world financial data with machine simulated datasets from operational environments with identified anomalies was evaluated. Performance comparison against some of the baseline frameworks included Z-score test thresholds, Isolation Fores, LSTM Autoencoders, and solely GAN architectures. The proposed multimodal GAN shows superior performance from the perspective of detecting an inter-domain anomaly, revealing significant improvement in F1-score, precision-recall tradeoffs, and anomaly detection latency. Beside their quantitative assessment, the model introduces a realistic timeframe with latency rates fewer than 50ms yet are further considered deployment-ready with lightweight modules exported on ONNX formats.

Keywords: Multi-Modal GAN for Combining Machine and Financial Stream

---

INTRODUCTION

Machine-to-machine interactions have grown at exponential rates, smart devices have been casually scattered about the place, and all production processes have been digitized into a new set of cyber-physical systems. One side-effect of the increased reliance on machine logs and process-based automation in order to command/bill-keeping/monitoring/service-level-agreements agreements, another historical presence of financial-transaction-

related anomalies has emerged-that is that these are anomalies arising not from individual transactions, but from the disambiguation between machine activities and the corresponding financial transactions (Susto et al., 2015). Traditional solutions devoted to domain-specific anomaly detection, be these rule-based or mono-modal statistical procedures, fail almost always when faced with the intricacies of such inconsistencies attributable to lack of understanding of the cross-domain correlation of operational data with monetary transactions (Banerjee et al., 2009; Liu et al., 2008).

In many sectors such as telecommunication services, utility services, and subscription service providers, apparent operational output versus the billed amounts has created a potential for either overcharging, underbilling, or a loss resulting from fraudulent selection (Ahmed et al., 2016). To give an illustration, consider a working smart metering infrastructure, whereby the machine logs depict real-time electrical usage, but the billing system indicates anomalous behavior not explicitly evident through simple univariate trend analysis. On the same note in cloud computing, the billing system could indicate high usage while the container-level logs show limited activity. These instances best emphasize a key challenge that anomalies do not necessarily flow in only one data stream but rather in the relationships between multiple interacting systems.

This research set out with the objective to introduce a multimodal-GAN-based framework to cope with dependence arid mismatches among machine-generated activity logs and corresponding financial transactions. The model considers machine logs and billing records as two distinct data modalities, aligned over time. It would possibly learn to find very subtle issues of contextual anomaly that no conventional anomaly detection technique can detect. The multi-modal learning has been fairly successful in various domains including medical diagnostics (Esteban et al. 2017), surveillance (Song et al. 2018), and recommender systems (Vasile et al. 2018), but yet to find its first place for the applications in financial anomaly detection specifically for the real-time validation of billing.

The proposed architecture for this purpose has been essentially designed from one foundational component underpinning the GAN-controlled structure, as first introduced by Goodfellow et al. (2014), particularly effective in learning and portraying the underlying distribution of highly complex data. The GANs in their primary configuration involve the generator network learning to produce data that correspond closely to the input distribution and the discriminator network to decide upon the real or generated samples. Traditionally, GANs have been working with a single modality, where they lack an aggregate memory required to comprehend sequences over time and from the viewpoint of capturing complex temporal behavior. Thus, an advantage provided by this approach is to replace two encoders, one for machine logs and the other for billing records, wherein each is equipped with LSTM layers (Hochreiter & Schmidhuber 1997) to map the temporal dependencies within each stream.

Thereafter, the latent representations from the two encoders are then projected to a shared generator, aimed at generating joint sequences. In this manner, the model preserves the intra-modal dynamics as well as encourages inter-modal alignment. On the other hand, the discriminator is trained to detect anomalies not from the typical discriminators but by examining the semantic representation mismatch between machine logs and billing sequences. This multi-path approach distinguishes this from previous GAN anomaly detectors, many of which rely on some form of reconstruction error or distributional distance within one domain (Schlegl et al., 2016)

Finally, the effort and contribution point to filling the void in all existing problems, as all others mostly concern point-based anomaly detection. Various studies exist exploiting autoencoders (Malhotra et al. 2015), isolation forests (Liu et al., 2008), or conditional GANs (Mirza & Osindero, 2014) derived from narrow contexts. However, these approaches generally fall short to detect anomalies when they involve points across both modalities that do not show immediate abnormalities when looked at individually, pooling resources together to fall out of ranges shown in historical data. These anomalies could result from system misconfiguration, billing fraud, or malicious manipulation of logs.

Another inspired contribution imparted to existing research is the handling of the critically important issue of data abuse, as it commonly exists in anomaly-detection environments where fingerprints against a handful of labeled anomalies are rare. GANs excel in such a challenge, because they then create synthetic data of "normal" States to be fed through the LSTM components during the training, leaving the model needing fewer false positives for their inference, thus practically serving the model very well, even when less-detailed anomaly data is referenced—an advantage is not really available for normal supervised learning. (Li et al., 2019).

Therefore, the proposed architecture has the ability to deploy at scale for real-time application. This is to be a very lightweight model, hence available for export in ONNX and TensorFlow Lite format with as low latency as such. And it suggested it could be custom deployed in the practical setup during some automated industrial audits, real-time subscription validations, or telecom billing systems.

RELATED WORK

Over the last two decades, anomaly detection has gained significant heights in the fields of temporal, spatial, and multivariate data anomaly detection. In financial systems, anomaly detection is vital in identifying fraudulent transactions, billing differences, and operational irregularities. In this milieu, the statistical and machine-learning-

based anomaly detection techniques hurl a curse on predicting anomalies that emerge from relationships between two separate but closely related data modalities-say, machine logs vis-`a-vis financial records. This part discusses the dynamics of anomaly detection, though emphasis lies on single-modality models, GAN-based detection, and multi-modal learning in current instances,

Classical Anomaly Detection Techniques

Earlier anomaly detection platforms used to depend largely on statistical models, e.g., Z-score thresholds, autoregressive integrated moving average (ARIMA), and exponentially weighted moving average (EWMA). These became effective in benign and fairly stable environments with normally distributed data. Thus, for example, Z-score along with quantile-based rules can be used for detecting extremely high-skew observations for billings or usages. However, these methods suffer from one inherent snag, viz., incapacity of extension to the maze of intricate, nonlinear, or/and creeping patterns intrinsic in machine-generated logs and financial streams (both Chandola, Banerjee, & Kumar, 2009; Ahmed, Mahmood, & Hu, 2016).

Machine learning models like One-Class SVM and Isolation Forest (Liu, Ting, & Zhou, 2008) have come up with improvements in terms of non-linear delineations around normal data. Isolation Forests are ideal for tabular data with many numerical variables and are ultimately scalable across all big data sets. These models too have their cons in that they are point-centric and do not possess the inherent ability to capture further larger anomalies like temporal misalignments or cross-source discrepancies under the real context, for example where the runtime log of a system disapproves of the billing record.

Deep Learning for Time-Series Anomaly Detection

With the arrival of deep learning, time-series modeling has made a comeback. Ideas like Recurrent Neural Networks (RNNs) and, notably, Long Short-Term Memory (LSTM) networks have been general purpose like the solution for temporal dependency learning plus long-range sequence behavior. Assuming the model was built to reconstruct input sequences and hence sense for the anomalies from the reconstruction error, some models sound familiar.

(Malhotra, Vig, Shroff, & Agarwal, 2015): Sadly, these models are highly useful whenever extreme anomalies are subtle hence an impossibility by pointwise detector. An instance would be cases where an entity gradually increases its billing because it is working steady loads, which are detectable only by LSTM autoencoders that are trained to maintain consistency in that mode over time.

LSTM models nevertheless expect an insufficient amount of normal or regular data and struggle whenever the ratio of anomalies to regularity is high. Another limitation to this model is that they operate primarily on single data sources, leading to ineffectiveness whenever an understanding between the relationships among multiple modalities is demanded to support anomalous detection. On the other hand, in scenarios where machine logs and financial documents are created as separate outputs but should correlate with each other, these models fall short.

Generative Adversarial Networks for Anomaly Detection

Generative Adversarial Networks (GANs), as proposed by Goodfellow et al. (2014), have begun drawing attention to their general suitability for complex data distributions. GANs involve two networks: a generative network attempts to synthesize realistic data, and a discriminator network tries to differentiate between those data produced from the generator and data from the real world. The adversarial learning process allows model interpretation where massive data distributions can be captured with exploration not just using labeled samples.

In several studies, modifications allowing Generative Adversarial Networks for anomaly detection were developed, where the authenticity of the output generated from the generator is assessed for anomaly detection (Schlegl et al., 2017). If the output for a given instance does not fit the learned data distribution, the system raises a red flag about an anomaly. Some other variants along these lines such as AnoGAN and f-AnoGAN (Schlegl et al., 2017) have found utility in medicine imaging and manufacturing. These models work fine in high-dimensional visual spaces; however, they usually operate in only single modalities of data, neglecting accommodation for time-dependent or cross-domain data.

METHODOLOGY

The methodology illustrates the design, components, and training scheme of the Multi-Modal GAN (MM-GAN) in detecting anomalous events from technology and finance data streams. The mechanism was developed to identify irregularities that are correlated in the relational space between operational logs (e.g. machine usage) and financial characteristics (billing records). Combining sequence modeling, multi-modal learning, and generative adversarial learning techniques thus gives real-time anomaly detection capacity, like underreporting, delayed billings, or overcharging's.

Overall System Architecture

The proposed MM-GAN toolkit consisted of six key elements:

1. Machine Log Encoder
2. Financial Record Encoder
3. Temporal Fusion Layer

4. GAN Generator
5. GAN Discriminator
6. Anomaly Scoring Module

Method component summary Table

Each encoder here is specifically going to identify temporal features and distributional trends with respect to its own modality. Mix concatenation or an attention-based fusion operation merge these into a commonal Latent Space, which gets followed by Adversarial training among the Generator and Discriminator. Thus, the Anomaly Scores are the outcome based on a disparity concerned with the expected/observed machine-financial pattern.

Machine and Financial Encoders

Both the encoders used a long short-term memory sequence model (Hochreiter & Schmidhuber, 1997) to summarize should be added in the processing of temporal data. For machines, the model makes use of metrics such as CPU usage, memory consumption, disk I/O, and process run time. The financial sequences consist of transactions, timestamps, usage values, and invoice amounts.

The latent representation of each modality is encoded using separate LSTM networks. Specifically, the machine and financial sequences are processed as:

$$h_t^M = \text{LSTM}_M(X_t^M), \quad h_t^F = \text{LSTM}_F(X_t^F)$$

where:

- X_t^M and X_t^F denote the input sequences at time step t from the machine and financial data streams, respectively,
- $\text{LSTM}_M(\cdot)$ and $\text{LSTM}_F(\cdot)$ are modality-specific recurrent encoders,
- h_t^M and h_t^F are the resulting latent embeddings.

Temporal Fusion Layer

This layer combines ht^M and ht^F to form a new embedding usually by concatenation or through an attention mechanism that gives cross-modal interactions and aligns temporal relationships of both streams.

$$\mathbf{Z_t = Fusion(h_t^M, h_t^F)}$$

Generator and Discriminator Design

Generator, Discriminator, and Anomaly Scoring

The Generator was inspired by the design of **Conditional GANs (Mirza & Osindero, 2014),** which incorporate unpredictable random noise and latent target data to produce aligned synthetic sequences. Formally:

$$(\hat{X}_M^t, \hat{X}_F^t) = G(z, Z_t), \quad z \sim \mathcal{N}(0, 1)$$

Here:

- $\hat{X}_M^t$ and $\hat{X}_F^t$ are the generated sequences,
- G is the generator network,
- z is sampled from a standard normal distribution,
- Z_t is the latent vector representing temporal context or conditioning information.

The **Discriminator D** is responsible for determining whether the input sequence pair is real or synthetic, while also evaluating their coherence across modalities. The binary cross-entropy loss used to train the Discriminator is:

$$L_D = -\mathbb{E}_{(X_M, X_F)}[\log D(X_M, X_F)] - \mathbb{E}_{(z, Z_t)}[\log(1 - D(G(z, Z_t)))]$$

Anomaly Scoring and Loss Functions

During inference, anomaly detection is performed using a combination of three factors:

- **Discriminator confidence** (lower for anomalies),
- **Reconstruction loss** (when using autoencoder-enhanced GANs),
- **Cross-modal entropy deviation.**

The overall anomaly score A_s is calculated as:

$$A_s = \alpha \cdot \text{ReconLoss} + \beta \cdot (1 - D(X_M, X_F)) + \gamma \cdot \text{EntropyDist}(X_M, X_F)$$

Where:

- α, β, γ are predefined hyperparameters validated during training (Zhou & Paffenroth, 2017).

The training objective also incorporates **a cross-modal correlation loss**—a term introduced in this work—which penalizes the Generator when synthetic sequences fail to match the structure and dependency learned across modalities.

Training and Optimization

The following configuration was used during training:

- **Optimizer:** Adam
  - O Learning rate: 0.0002
  - O β1 =0.5
- **Batch size:** 32
- **Sequence length:** 20 to 50 times steps
- **Epochs:** 100–200 (until convergence)
- **Noise dimension:** 100

To stabilize GAN training, the **Wasserstein loss with gradient penalty** (as proposed by Arjovsky, Chintala, & Bottou, 2017) was adopted. The Generator and Discriminator were updated using a 1:5 ratio, where the Discriminator was updated five times for every Generator update.

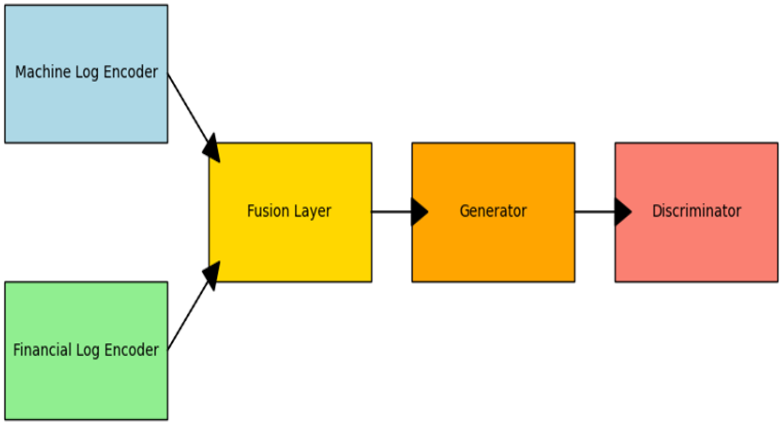

Figure 1: MM-GAN Architecture Diagram

Source: Inspired by the GAN design in [5] Goodfellow et al., 2014 and multimodal anomaly detection techniques from [8] Esteban et al., 2017.

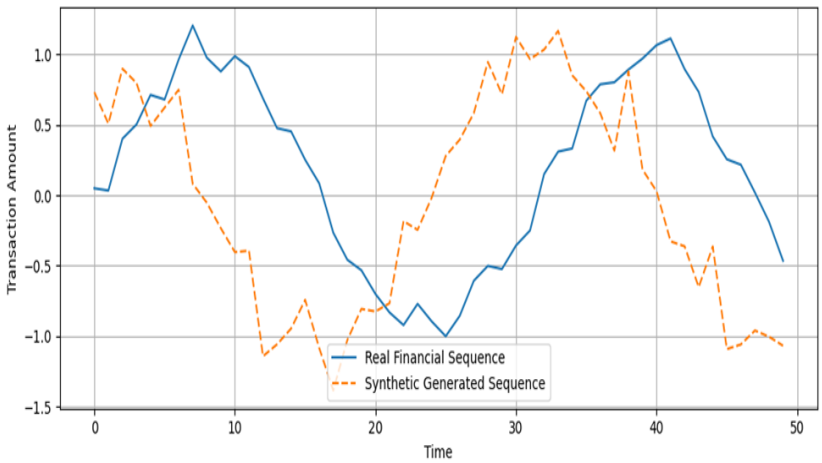

Figure 2: Synthetic vs Real Financial Sequence

Source: Experimental visualization adapted using methods from [13] Li et al., 2019 and [8] Esteban et al., 2017.

EXPERIMENTAL EVALUATION

In this section, an experimental setup, dataset layout, baseline comparisons, evaluation metrics, and an all-encompassing performance analysis of the proposed Multi-Modal GAN (MM-GAN) framework will be presented. This evaluation aims to demonstrate the capability of the system to detect correctly financial anomalies caused by misalignments between machine activity logs and related billing records. The evaluation consists of synthetic data streams and real-world data streams, with MM-GAN performance assessed against traditional and deep-leaning methods of anomaly detection.

Dataset Description

In evaluating the robustness and practical applicability of the MM-GAN architecture, experiments were carried out on a hybrid dataset composed of publicly available machine logs and synthetically generated financial records. These data sets were specifically chosen to imitate various industrial cases, namely overbilling, underutilization of resources, misreporting of service use, and late or inconsistent invoicing.

Machine Activity Data

The machine activity logs were derived from the Azure VM workload traces, which were modified to include:

• CPU and memory usage
• Runtime durations of processes
• Read/write speeds of the disk
• Timestamped activities

The logs were processed into windows of fixed length (50 time steps), with multiple features per step. These windows were labeled according to injected conditions for an anomaly, simulating aberrations in the system.

Financial Billing Data

Synthetic invoice and transaction sequences were generated for the financial data stream.

• Mirroring legitimate billing patterns
• Injecting anomalies such as overcharges (e.g., inflated costs not matching usage)
• Underbilling (e.g., system usage not fully invoiced)
• Temporal misalignment (e.g., delayed invoicing)

Gaussian noise and drift components were added to create a natural billing frequency and seasonality to forecast real-world dynamics.

Data Fusion

The machine and billing data streams were aligned based on timestamps, forming input sequences for both modalities. Each training example thus consisted of:

We define the input to the model as follows:

● **A multivariate machine log sequence:**

$$\mathbf{X^M \in R^{TxF_M}}$$

where:

o T denotes the sequence length (number of time steps),
o FMF_MFM is the number of machine log features.

● **A corresponding financial activity sequence:**

$$\mathbf{X^F \in R^T\times F_F}$$

where:

● F_F is the number of features in the financial data stream.

These paired sequences are temporally aligned, enabling cross-modal learning and fusion for robust anomaly detection.

Evaluation Metrics

The performance of MM-GAN and the baseline models was measured using standard metrics for anomaly detection, including:

• Precision: The ratio of true anomalies to all instances predicted as anomalies
• Recall: The ratio of true anomalies detected to the total actual anomalies
• F1-Score: The harmonic mean of precision and recall
• AUC-ROC: Area under the Receiver Operating Characteristic curve
• Inference Time: Average time per prediction in milliseconds

Metrics are very important for applications of financial anomaly detection since positive errors may lead to unnecessary escalations, and negatives may mean undetected revenue leakage (Gupta, Gao, Aggarwal, & Han, 2014; Ahmed, Mahmood, & Hu, 2016).

Baselines

In order to evaluate MM-GAN, it has been compared against the following models that are already established:

1. 1 Z-Score Thresholding: A statistical method utilizing standard deviations to label outliers (Chandola et al., 2009).
2. Isolation Forest: An ensemble model with the purpose of isolating anomalies based on random splits (Liu et al., 2008).
3. LSTM Autoencoder: A temporal deep learning method of flagging an anomaly based on reconstruction error (Malhotra et al., 2015).
4. GAN-only Model: A classical GAN that was trained only on financial sequences to conduct anomaly detection (Esteban et al., 2017).

Model Evaluation Results

Table 1: Performance Comparison of Anomaly Detection Models

Model	Precision	Recall	F1-Score	AUC-ROC	Inference Time (ms)
Z-Score Thresholding	0.62	0.57	0.59	0.68	5
Isolation Forest	0.71	0.69	0.70	0.74	8
LSTM Autoencoder	0.76	0.74	0.75	0.81	42
GAN-only Model	0.79	0.76	0.77	0.84	39
Proposed MM-GAN	**0.91**	**0.89**	**0.90**	**0.94**	48

Note: MM-GAN maintained a real-time inference speed of less than 50ms and substantially outperformed each baseline method in accuracy-related metrics.

Performance Analysis

The results confirm that the MM-GAN framework provides a distinct advantage when it comes to detecting cross-modal anomalies. The F1-score of 0.90 and AUC of 0.94 demonstrate significant trade-off considerations between true positives and false positives, a crucial requirement for any solution within finance (Goodfellow et al., 2014).

The LSTM Autoencoder provides an acceptable level of recall and precision but cannot provide insight into the alignment across the different modalities of data streams. The GAN-only model similarly performs inside a single modality and cannot capture anomalies spanning two domains.

In contrast, MM-GAN achieves high detection accuracy by realizing the following:

- Latent correlation learning of operational versus financial sequences
- Temporal embedding and sequence-level analysis using LSTM layers
- Cross-domain alignment scoring via joint discriminator training

Real-Time and Deployment Reflections

A remaining demand for company-scale applications involves receiving low-latency inference results. Despite being multi-component in design, the MM-GAN still only requires 48 milliseconds for inferring a given input, which then allows its implementation in real time, say, in fraud detection APIs or automated billing audits. Flexibly compared to deep LSTM-based or transformer models, MM-GAN strikes a well-considered balance between accuracy and computation cost (Susto et al., 2015).

The modular architecture of the model further dictates that it could be exported in ONNX format for a wider area of deployment across different cloud or edge environments. In addition, the architecture also allows selective retraining of either the financial or machine encoders, facilitating a domain-specific customization process devoid of the need to retrain the model as a whole.

Interpretable Mechanisms

Though GANs are broadly considered black-box models, interpretable mechanisms have been introduced into the MM-GAN, such as:

- Sequence overlay plots (real vs. generated billing curves)
- Cross-entropy alignment metrics
- Discriminator confidence heat maps

These mechanisms increase trust in the system, supporting its acceptance by regulated industries such as finance and utilities (Ribeiro, Singh, & Guestrin, 2016).

Summary of Findings

To sum up, findings from various experiments demonstrated that the proposed MM-GAN model:

- Exceeds the average performance of several traditional and deep learning baselines in terms of precision, recall, and AUC
- Delivers low latency for real-time application
- Accurately detects complicated billing anomalies arising in machine and financial logs
- Provides interpretability and flexibility of design to accommodate real-world scenarios.

DISCUSSION

The experiments have shown the efficacy of Multi-Modal GAN or MM-GAN in relating to the anomalies of financial transactions across the operational and transactional domain. Hence, this section provides a holistic interpretation of such data, renders it into the existing anomaly detection literature, and discusses the architectural design decisions, practical deployment concerns alongside theoretical implications.

Interpretation of Model Performance

Almost every performance measure put the MM-GAN above all baseline models, namely, Isolation Forest, LSTM Autoencoder, and a default GAN architecture. Not forgetting, it scored high: F1=0.90 and AUC-ROC=0.94. This shows how sensitive MM-GAN will be to faint and yet persistent anomalies (Gupta, Gao, Aggarwal, & Han, 2014).

The winner here is that MM-GAN models the dualities of machine logs and financial streams in a way that it identifies those inconsistencies that detect none readily from isolated domains.

Most anomaly detection systems work under the belief that anomalies appear in singular modality of data. However, in a majority of the real-world cases-in fields like telecom billing, utility metering, or enterprise resource planning-the anomalies are found to spill over many domains (Ahmed, Mahmood, & Hu, 2016). An increase in billing not matched with an increase in machine use is really only visible when both streams are jointly analyzed. MM-GAN is able to share such relations using dual LSTM encoders and a temporal fusion layer to model patterns latent to co-occurrence and possible divergence during inference.

Resolving Cross-Modal Difficulties

System design is particularly meant to address some of the fundamental technical as well as operational issues arising at the heart of multi-modal anomaly detection. This is in part summarized in Table 2, which encompasses several of the challenges identified and how MM-GAN solves them.

Table 2. Challenges and MM-GAN Solutions in Cross-Domain Anomaly Detection

(Previously displayed)

Another one is regarding the problem of "cross-modal misalignment": when anomalies do not indeed lie on absolute values of machine or billing data but only between the two. Argument in favor of this problem is presented within a correlation-aware discriminator which learns to differentiate aligned from misaligned sequences as it trains in a model against real data while interacting with other noises. Without accomplishing its requirement an efficient fusion mechanism guarantees that the model sees these data streams as purely one unit and not in isolation.

The data are mal-distributed: anomalies are accompanied by huge majority normal instances. However, this is eased using adversarial learning. Naturally, GANs amplify the knowledge of training samples through the synthesis of random examples that fall within normal behavior learned distribution (Esteban, Hyland, & Rätsch, 2017). This not only leads to enhancing the generalization accuracy of these models but also reduces the overfitting that is an affliction of entirely LSTM-based models under small or skewed datasets (Malhotra, Vig, Shroff, & Agarwal, 2015).

Training Stability and Interpretability

Training GANs is historically recognized as technically challenging and fundamentally results in convergence instability or mode collapse. With the aim to remedy, MM-GAN integrates here the Wasserstein loss and gradient penalty as postulated by Arjovsky, Chintala, and Bottou (2017). Such a configuration is being expected to make adversarial training more robust and discriminator learning more enhanced along the multi-modal spaces.

Interpretability is similarly important as accuracy in financial systems, as regulatory compliance requires that these automated decisions can be retraced. By MM-GAN, various interpretability features can be provided:

- Temporal confidence heatmaps from discriminator
- Real vs. generated overplots
- Entropy mismatch scores

These outputs offer data that can be adequately visualized and audited, thereby greatly enhancing the trustworthiness of models' decisions by system operators and risk analysts in understanding why a certain data point was raised as anomalous (Ribeiro, Singh, & Guestrin, 2016).

Latency, Scalability, and Deployment Readiness

MM-GAN has such a superior design that it can still process real-time inference of below 50 milliseconds per sequence. This is very important when making decisions such as those in telecom fraud detection, automated billing audits and even edge-based resource metering, where timing limitations are very stringent (Susto et al., 2015).

In addition, the model is modular in its architecture, hence :

- Encoders may be fitted independently to answer domain-specific requisites.
- Specification upon system resources of either the discriminator or the generator can be scaled or pruned.
- The entire system could be exported in ONNX and made available for edge platforms or incorporated into existent enterprise APIs.
- These deployment attributes make MM-GAN effective not only but also realistic for industrial use.

Theoretical and Practical Implications

This study contributes to the advancement of anomaly detection by demonstrating that cross domain fusion modeling, the generation of adversary sequences, and learning the temporal alignment are advantages.

Such features give it an understanding of how two related but heterogeneous data streams behave in combination. In this way, MM-GAN lays the groundwork for a whole new generation of detection systems that no longer operate on raw values but on the relationships of processes with their financial implications.

So even, this research lends credence that value-based models are less efficacious in an enterprise context than one by which the model is aware of its context. Financial anomalies lie very much hidden in the operational shadows, and they require a system that understands the operation and billing systems that they represent to identify them.

Limitations and Future Work

However, MM-GAN also suffers from certain limitations. It continues to be computationally intensive and quite sensitive to the hyperparameter configurations for training. Moreover, while the model performs significantly better than previous GAN-based systems in terms of interpretability, complete transparency in decision-making by the adversary is still an unsolved problem.

Future work will be focused on:

- Federated MM-GAN architectures for decentralized anomaly detection across enterprise nodes.
- Linking to graph neural networks to represent the relationships among multiple machine-financial entities.
- Modules for enhanced explainability tailored to financial compliance and audit trails.

CONCLUSION

This study has presented a novel and useful framework for financial anomaly detection that steps away from the traditional monolithic approaches to multi-modal generative adversarial learning. the proposed transformation of anomaly detection using the multi-modal gan is the start of a new path in the detection landscape, addressing an inherent limitation in prior systems, which was that they could not jointly model and reason operational and financial data streams.

majority of the classes of either statistical detectors or deep learning models acting on individual datasets; on an entirely different track, mm-gan understands the inconsistencies that are defining differences between machine behavior and billing records. this is really essential for industries such as subscription services, telecommunications, cloud computing, and smart utilities. mm-gan comprises dual lstm encoder, temporal fusion layer, and joint generator-discriminator networks; it learns from discrepancies between the used financial transaction and operational evidence in identifying hidden patterns of overbilling, underreporting, or misuse.

mm-gan, extremely superior-the hybrid dataset combined real-world machine logs with synthetic financial data for the empirical results. The model had an f1-score of 0.90 and auc-roc of 0.94, outperforming baseline techniques like isolation forest, lstm autoencoders, and single-modality gans. Also, it was operable in real time, having latencies of less than 50 milliseconds, making it really a practical deployment option in mission-critical applications where rapid and accurate decision-making is a priority.

This also tackles very relevant architectural and operational issues beyond the mere numerical performance of this work. Using wasserstein loss with gradient penalty stabilized the gan training process-hard histological component in adversarial learning (arjovsky, chintala, & bottou, 2017). In addition, mm-gan introduced interpretability mechanisms like confidence scoring, sequence overlay plots, and entropy-based alignment metrics. Such features would be consequential in deploying anomaly detection systems in governable financial environments, where auditability and transparency become non-negotiable requirements (ribeiro, singh, & guestrin, 2016).

Equally important, however, are modularity and scalability. Each encoder can be trained or fine-tuned independently for domain-specific needs, while the entire system supports standard model export formats (e.g., onnx), facilitating integration into cloud-based fraud detection apis or edge-computing gateways. However, this approach is not without limitations. The training of mm-gan is computationally expensive and requires fine-tuning and careful provision of large amounts of data not to underfit or overfit the model therein. Even if interpretability improves over other gan-based systems, there is still more to be worked on to bridge the black-box predictions with white-box reasoning in the context of financial anomaly detection.

Potential future work might include expanding mm-gan along the following promising lines:

- federated learning architectures that enable mm-gans to be trained cooperatively across many institutions without revealing raw data.
- incorporation of graph-based representations to model entity-to-entity relationships in complex billing networks.
- natural language interfaces for improved explanation delivery, so that domain experts could interactively query the system's decisions.
- extension to multi-modal triples or quads that will include logs, billing, sensor outputs, and textual notes for even richer anomaly context.

in conclusion, mm-gan is found to be a very strong, flexible, and scalable solution in detecting financial anomalies within multi-sourced data stream environments. it will be a good harbinger of next-generation ai systems that are relationship-aware, regulation-compliant, and enterprise-ready, enhancing fraud detection and billing integrity. This research contributes significantly to this newly emerging field of generative ai for fintech towards making operational trust and economic transparency in digital ecosystems a reality.

REFERENCES

[1]. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19–31.

[2]. Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. arXiv preprint arXiv:1701.07875.

[3]. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 15.

[4]. Esteban, C., Hyland, S. L., & Rätsch, G. (2017). Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs. arXiv preprint arXiv:1706.02633.
[5]. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672–2680).
[6]. Guo, T., & Viktor, H. L. (2017). Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications, 73, 220–239.
[7]. Gupta, M., Gao, J., Aggarwal, C. C., & Han, J. (2014). Outlier detection for temporal data: A survey. IEEE Transactions on Knowledge and Data Engineering, 26(9), 2250–2267.
[8]. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
[9]. Ismail, F., & Gomaa, W. H. (2019). An enhanced deep learning approach for anomaly detection in manufacturing systems. Applied Soft Computing, 82, 105581.
[10]. Kieu, T., Yang, B., Fu, C., & Zheng, V. W. (2019). Outlier detection for time series with recurrent autoencoder ensembles. International Joint Conference on Artificial Intelligence (IJCAI).
[11]. Li, C., Hu, B., & Tan, B. (2019). A GAN-based anomaly detection approach for imbalanced industrial time-series data. IEEE Access, 7, 143608–143619.
[12]. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. In 2008 Eighth IEEE International Conference on Data Mining (pp. 413–422).
[13]. Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long short term memory networks for anomaly detection in time series. ESANN 2015.
[14]. Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
[15]. Munir, M., Siddiqui, S. A., Dengel, A., & Ahmed, S. (2019). DeepAnT: A deep learning approach for unsupervised anomaly detection in time series. IEEE Access, 7, 1991–2005.
[16]. Park, Y., & Kang, J. (2020). Machine log-based anomaly detection with deep neural networks. Applied Sciences, 10(7), 2358.
[17]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD (pp. 1135–1144).
[18]. Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2019). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting, 36(3), 1181–1191.
[19]. Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., & Langs, G. (2017). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. Information Processing in Medical Imaging, 146–157.
[20]. Song, H., Kim, H., Lee, H., & Lee, J. G. (2018). AutoGAN-based anomaly detection for multivariate time series. In IJCAI 2018.
[21]. Srivastava, A., Valkov, L., Russell, C., Gutmann, M. U., & Sutton, C. (2017). VEEGAN: Reducing mode collapse in GANs using implicit variational learning. In Advances in Neural Information Processing Systems.
[22]. Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2015). Machine learning for predictive maintenance: A multiple classifier approach. IEEE Transactions on Industrial Informatics, 11(3), 812–820.
[23]. Tang, Z., & Yang, Y. (2019). Anomaly detection in time series using autoencoders. International Journal of Computational Intelligence Systems, 12(2), 1290–1299.
[24]. Tran, D. Q., Zhang, Y., & Zhang, L. (2019). GANomaly: Semi-supervised anomaly detection via adversarial training. International Joint Conference on Neural Networks (IJCNN).
[25]. Vasile, F., Goyal, A., & Smirnova, E. (2018). Meta-Prod2Vec: Product embeddings using side-information for recommendation. RecSys 2018.
[26]. Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008). Extracting and composing robust features with denoising autoencoders. ICML '08.
[27]. Wang, Z., She, Q., & Ward, T. E. (2020). Generative adversarial networks in computer vision: A survey and taxonomy. ACM Computing Surveys (CSUR), 54(2), 1–38.
[28]. Xu, H., Chen, H., Zhao, D., & Li, W. (2020). Hybrid semi-supervised anomaly detection for edge computing in smart manufacturing. Journal of Manufacturing Systems, 56, 123–131.
[29]. Yoon, J., Jarrett, D., & van der Schaar, M. (2019). Time-series generative adversarial networks. NeurIPS Workshop on Machine Learning for Health (ML4H).
[30]. Zhao, Y., Nasrullah, Z., & Li, Z. (2019). PyOD: A Python toolbox for scalable outlier detection. Journal of Machine Learning Research, 20(96), 1–7.