



SELF-SUPERVISED KNOWLEDGE REPRESENTATION FOR RARE FRAUD AND OPERATIONAL FAILURE DETECTION IN MULTI-CHANNEL PAYMENT SYSTEMS

Boumedyen Shannaq¹, Nour Eldin Elshaiekh², Basel Bani-Ismail³, Marwan Alshar'e⁴, Oualid Ali⁵

¹Management Information system department, College of Business, University of Buraimi, Sultanate of Oman

²Department of Information Studies, Faculty of Arts and Social Sciences, Sultan Qaboos University

³Department of Software Engineering, Faculty of Information Technology, Applied Science Private University, Amman, Jordan

⁴Faculty of Computing and IT, Sohar University, Sohar, Oman

⁵Department of Computer Science, College of Arts and Science, Applied Science University, Manama, Kingdom of Bahrain

¹<http://orcid.org/0000-0001-5867-3986>, ²<https://orcid.org/0000-0001-7465-2675>, ³<https://orcid.org/0000-0002-1851-315X>

⁴<https://orcid.org/0000-0001-5187-4902>, ⁵<https://orcid.org/0009-0001-1433-017X>

E-mail: *boumedyen@uob.edu.om.

ARTICLE INFO

Article History

Received on: March 03, 2026

Reviewed on: April 02, 2026

Accepted on: May 01, 2026

Published on: July 31, 2026

Keywords:

Self-Supervised Learning,
Fraud Detection,
Knowledge Representation,
Payment Systems,
Financial Information Systems,
Imbalanced Learning.

ABSTRACT

In modern high-volume payment systems, detecting fraud is still essentially confined by abhorrent class imbalance, changing transaction patterns, and lack of dependably labelled fraud occurrences. The current research questions the issue of whether self-supervised learning (SSL) can add to the extraction of the knowledge related to fraud in comparison with the capability of strong supervised baselines in the multi-channel payment setting. Using a real world banking dataset of over 13.3 million transactions in the 2010-2019 period, we perform an extensive analysis, including supervised machine learning, anomaly-based SSL, and methods of integrating knowledge into machine learning strategies. Gradient-boosting models are able to build a strong base ($F1 = 0.86$, $ROC\text{-}auc = 0.99$) that suggests that the trained model has a near-saturation discriminative ability that is solely based on tabular transaction characteristics. We show that naive, generic, SSL-based anomaly detectors lead to reduced precision, and task-adapted representations of supervised models, stacked with task-adapted representations, can increase fraud recall by up to 4.9 with a small $F1$ increase (+0.6). However, with strict operationally imposed conditions of accuracy ≥ 0.90 , the added benefits of the use of SSL are not experienced, highlighting inherent thresholds of representation based improvement. Such results enhance a knowledge based perspective of when self-supervised representations add value to the decision making and when supervised models have already acquired adequate information about fraud meaning thereby guiding the design of financial fraud knowledge-management models in a robust way.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The rapid digitization of payment ecosystems has radically transformed how financial institutions should create, handle and operationalize knowledge towards preventive fraud [1]. Modern banking systems are now handling millions of heterogeneous transactions over various channels swipe chip and online, in which fraudulent activity is an exceptionally rare but high-impact event. This discrepancy is further exacerbated by the slow or partial labeling, which can be very problematic to the classical supervised learning methods that assume stable distributions and rich labeled examples [2], [3]. Based on this, fraud analytics is thus being redefined as a knowledge representation and discovery problem in which the ability to model normality at scale is becoming a crucial component as much as the ability to identify familiar fraud patterns [4]. The recent developments in the sphere of machine learning highlight the shortcomings of a strictly supervised model in the context of changing user behavior [5], [6], [7] channel migration and operational noise.

Even very good classifiers can become untrustworthy as fraud schemes change or as innocent operational failures, including technical failures or authentication errors, mimic atypical behavior [8]. This has spawned an interest in self-supervised learning (SSL), that learns the latent representations by using unlabeled data, so that systems can learn the consistent structural knowledge by using the transaction attributes, temporal interactions, and categorical interactions[9], [10]. A systematic literature review on cyber threat was investigated by [11] on cyber threat Intelligence for organizational cybersecurity resilience. In the literature on knowledge management, the concept of SSL is taking on a new form as a tool of extracting the reusable organizational knowledge and not just a preprocessing step. [12] Improves behavioral security analytics by doing user behavior identification, and it also complements the self-supervised framework that's being proposed, for spotting rare fraud events and operational breakdowns in multi-channel payment systems. [13] reviews self-supervised recommender systems, showing how they reduce data sparsity and label dependence while improving user and item representations across recommendation tasks. [14] provides a thorough analysis of intrusion detection systems, discussing the latest developments in artificial intelligence and machine learning-based IDS and various deployment scenarios, major challenges, and future research opportunities in advancing network cybersecurity. However, there is growing evidence that SSL does not consistently perform well on strong gradient-boosting baselines on tabular financial data especially when the data has operations that already capture most of the information that influences the decision being made in Generative hybrid models [15]. This leads to an important research question in information and knowledge management: when does representation learning provide any new knowledge, and when does it saturate the decision structures already provided? The solution to this question is to conduct strict comparative analysis on the basis of realistic deployment policies, among which there are threshold governance and precision-constrained decision rules[16]. This paper aims to explore a self-managed anomaly detection model on an eleven-year, multi-channel payment system containing more than 13 million transactions. The study assesses the effect of representation learning on fraud knowledge extraction by systematically comparing supervised baselines, SSL-enhanced models, and hybrid stacking strategies to obtain a better understanding of the impact of representation learning on fraud knowledge extraction in extreme class imbalance. The contribution does not therefore claim the universal superiority of SSL but explains the operational value, constraints and knowledge implications of fraud analytics against the increasingly vital issue of financial information systems in the modern era, the use of machine learning techniques in Financial fraud detection was reviewed by [17], and the Exploring biometric technology was investigated by [18].

II. RELATED WORK

In recent years, the focus in fraud analytics has gradually moved from traditional label-intensive supervised learning to knowledge-rich representation learning methods that are more resilient to extreme class imbalance, delayed label availability and the ever-changing nature of payment transaction characteristics [19-23]. In this context, self-supervised learning (SSL) has proven to be a promising learning framework for extracting informative latent features from massive amounts of tabular data with no labels. Modern SSL models aim much more than preprocessing; they define stable semantic structures from diverse attributes and payment channels, and models are able to capture these subtle anomalies and characterize the normal behavioral patterns [24], [25], [26]. Further, recent surveys confirm that SSL performs best when using strong downstream predictive models, especially when pretext tasks are carefully crafted to be appropriate to the corruption mechanisms in the domain, such as feature masking, temporal perturbation, and swapping of categorical values, rather than generic data augmentation tasks [26], [27], [28], [29]. Even with these progressions, financial fraud detection pipelines still heavily require Explainable Artificial Intelligence (XAI) to improve the transparency of the models and to enable trustworthy financial decision making in high-risk settings. For instance, in mobile financial fraud, explainable AI has proven to be effective [30] and, in the field of financial forecasting, AI has shown excellent predictive power at the firm level [31].

These methods are increasingly adopting SSL as a side-component of representation learning, but it appears that discriminative representation learning is superior to generative representation learning for discriminating between fraudulent transactions and operational failures. Especially, diffusion-based models and autoencoders are no longer judged only by their reconstruction performance, but also by their discriminative performance [15-32]. Moreover, it was proven that machine learning technologies can be used to successfully support predictive decision making and intelligent anomaly prediction in complex financial environments, this is an important methodological basis that has been established for the payment risk assessment in modern payment systems [32]. To address this, anomaly detection algorithms based on anomaly-discriminating representations, in addition to reconstruction errors, have become more prevalent in recent years [33]. Conventional generative detectors, which frequently reconstruct fake observations with high accuracy, however, decrease the sensitivity of the detector to anomalies [34] that has been a major limitation of such models. Self-supervised diffusion models, for instance, have been specifically designed with discriminative objectives to address this issue. Meanwhile, graph-based fraud detection works by building a dynamic and relational graph of payment relationships, allowing the discovery of coordinated fraud activities which would be hard to capture using transaction-level features alone [35].

The need for explainable AI in enhancing the transparency and interpretability of fraud detection systems has also been emphasized in complementary research, fostering user trust and aiding in informed operational decisions, a systematic review have been investigated by [36]. However, the feasibility of implementing graph-based solutions is often hindered by computational complexity, the availability of graphs, and the scalability of the solutions in practice, especially in industrial payment systems. Many practical financial applications are still using tabular learning frameworks with some relational attributes and not complete graph-based approaches [37] [38]. Previous empirical research has also shown that machine learning is effective in assessing organizations' risk and predicting anomalies [39], which also validates the potential application of self-supervised representation learning in the detection of rare fraud events and operational failures in multi-channel payment setting. [40] offers a contemporary overview of the state-of-the-art of deep learning by exploring key recent developments in architectures, innovative training methods, various practical applications, and promising research avenues to further push the boundaries of AI. In other security areas, self-supervised learning for feature learning has been shown to be beneficial and has been successfully combined with gradient-boosting methods for detecting intrusions in IoT networks [41].

Moreover, recent evaluation studies have revealed that the widely cited ROC-AUC metric could give an overly favourable evaluation on highly imbalanced datasets, for which the practical deployment requires to attain acceptable precision-recall trade-offs. For financial applications, PR-AUC, combined with operational threshold selection strategies (such as maximizing the F1 score or optimizing for precision) have gained increased significance as a measure of the performance of rare-event detection systems in real-world applications [42] [43]. Although these are crucial advances, there are also a number of limitations with the current studies. Most methods use auxiliary SSL modules and do not fully utilize self-supervised knowledge representation for discriminative learning, and few frameworks explicitly consider the identification of rare fraud events and operational failures of heterogeneous payment channels simultaneously. Furthermore, there is a lack of works focusing on combining powerful self-supervised representation learning and providing explainable decision support with operationally relevant evaluation criteria. The main research gaps found in the literature are summarized in Table I, which also inspires the proposed framework for Rare Fraud and Operational Failure Detection in Multi-Channel Payment Systems: Self-Supervised Knowledge Representation. Table I summarizes the main research gaps found in the literature and motivates the proposed framework for Rare Fraud and Operational Failure Detection in Multi-Channel Payment Systems: Self-Supervised Knowledge Representation.

Table 1: Research Gap Summary Table.

Gap observed in recent work	Why it matters for multi-channel fraud/ops	How this work addresses it
SSL often reported without policy-level thresholding	Banks must commit to precision targets; “better AUC” may not reduce false alerts	Report best-F1 and precision $\geq$ 0.90 operating points
Generative/AE anomaly models blur fraud vs. benign declines	Operational failures create hard negatives; mislabeling harms knowledge actions	Treat failures as structured “near-anomalies” and tune thresholds using fraud labels
SSL gains saturate against strong GBDT baselines	Tabular GBDT already encodes many stable rules; SSL must add new knowledge not redundancy	Use SSL primarily to improve minority recall under constraint, not to “replace” LightGBM
Graph methods are powerful but costly to deploy at scale	Real-time graph construction/updates are heavy in practice	Keep deployment lightweight: embeddings + anomaly scores + calibrated classifier

Source: Authors, (2026).

III. METHODOLOGY

This work presents a knowledge based detection of fraud model where controlled learning is communally coupled with self-managed representative learning. Thus eliminating the twin tragedies of profoundly unequal class distribution and limited quantity of marked instances that are widespread in massive payment ecosystems. The main contribution is that the normative patterns of transaction are identified with the help of the self-supervised embeddings. And further incorporated into the high-performing supervised ensemble through stacking, instead of replacing the latter. This trick is effective in that the anomaly knowledge can be used to complement the fraud recall at the expense of the powerful discriminative ability of gradient-boosting classifiers. Besides, the method analyzes decision strength at a high operational threshold (such as having accuracy of at least 0.90) hence reflecting the requirement of modern fraud-management procedures.

III.1 DESCRIPTION AND COMPLEXITY OF DATA SETS

The collected dataset is 13.3M verified banking transactions between 2010 and 2019 [44], with only 13,332 cases (0.1 percent) of them (so-called confirmed fraud) representing an incredibly biased learning task. All the records itself contain heterogenous features such as payment amount, time of transaction, payment channel (swipe, chip, online) and merchant category codes (MCC) as well as geographic indicators (e.g., invalid PIN, technical glitches). The scale, diversity, time span and scarcity of fraud incidents make this corpus unavoidably complex, which requires advanced representation learning algorithms and rigorous evaluation procedures. Proposed Methodology and Experimental Setting

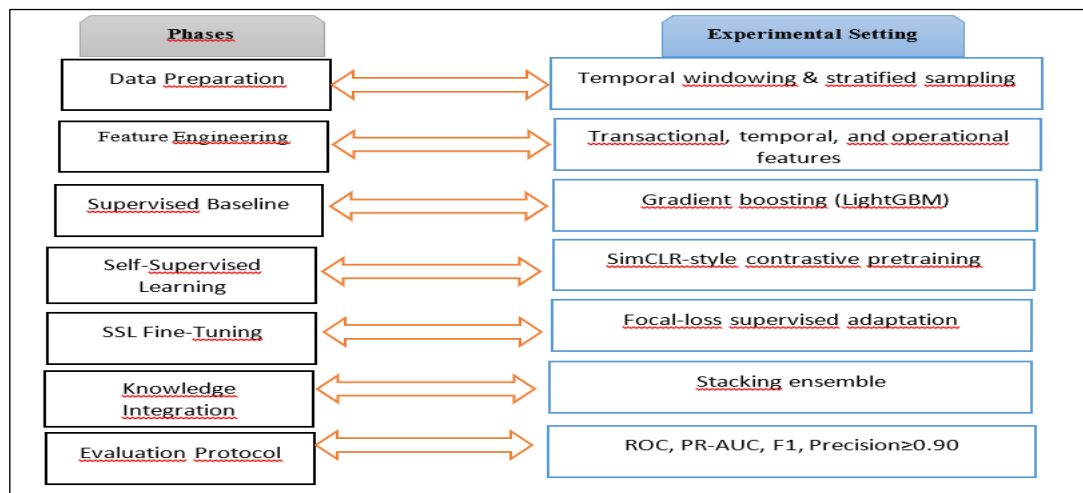


Figure 1: Proposed Methodology and Experimental Setting.

Source: Authors, (2026).

The experimental process starts with data preparation whereby the temporal windows that are rich in fraud are automatically identified and then balanced, with the use of stratified sampling and controlled negative sampling to maintain the authentic patterns of fraud. The following step is feature engineering, which includes transactional, temporal and operational variables, numerical features (amount and time of transaction encoded) and categorical features (channel, merchant category code (MCC), state) transformed, and operational failure-type signals are included to learn rich multichannel behaviour. A monitored base model is then trained on the gradient-boosting (LightGBM), defining a performance floor on detection of frauds in a class learning environment. In addition to this, self-supervised learning is proposed with the help of a contrastive pretraining style based on SimCLR, in which only normal transactions are used to learn latent representations of typical payment behaviour, without any reference to labels of fraud. The resulting embeddings are then optimized during the fine-tuning phase by using supervised learning with focal loss, and, therefore, aligning the representations with fraud-detection goals in highly imbalanced conditions. The system also incorporates the knowledge of both supervised and self-supervised models in stacking ensemble and thus, combines complementary decision cues. Lastly, ROC, PRb-AUC, F1-score, and operation gauges (e.g. keeping precision 0.90 and above) are strictly evaluated as model performance metrics in both statistical strength and practical application effectiveness.

III.2 METHODOLOGICAL NOVELTY

In this attempt, we introduce a knowledge-intensive approach to detecting frauds synthesising the self-supervised models of the normative transactional behaviour with the state-of-the-art supervised classifiers through decision-level stacking. Contrary to the antecedent methodologies that replace supervised learning directly, our method enhances the fraud acumen in the context of the lack of labelled data, thus improving the recall without compromising the reliability of the operations. In addition, the structure examines the strength of the decisions with very high levels of precision requirements, where there is agreement between the behaviour of the model and the practical policies of managing fraud.

IV. RESULTS AND ANALYSIS OF THE EXPERIMENT

The experimental validation of the proposed framework over a series of five experiments outlining a route over the traditional supervised baselines to the integration of the self-supervised representations and finally the modulation of operational cut-off points. Representation of results is in terms of the quality a capital of mental essence it adds, accuracy of soundness measurement and delicacy of the text of choice, as opposed to solitary performance measurements.

IV.1 BASELINE PERFORMANCE (EXP1) UNDER SUPERVISION

Table 2: Supervised Model Comparison (Exp1).

model	precision	recall	f1	roc_auc	pr_auc
LightGBM	0.896619	0.931745	0.913844	0.990289	0.97043
XGBoost	0.883245	0.941034	0.911224	0.989689	0.970068
RandomForest	0.934838	0.886511	0.910033	0.987088	0.962538
LogisticRegression	0.841756	0.921648	0.879892	0.979046	0.942459

Source: Authors, (2026).

Table 2 delivers a comparative analysis of the canonical supervised algorithms, namely, Logistic Regression, Random Forest, XGBoost, and LightGBM, applied on the measures of F1-score, ROC-AUC and PR-AUC. LightGBM achieves the best overall performance (F1 $\approx$ 0.91, ROC-AUC $\approx$ 0.99) which forms a high-performance standard in detecting fraud in a situation where only tabular transaction features are accessible. These results endorse the notion that modern gradient-boosting algorithms are able to infer much information about frauds on a transactional level, creating a solid foundation on which future developments can be built.

IV.2 SELF-SUPERVISED (EXP2) ANOMALY DETECTION

Table 3: SSL Anomaly Detection Results (Exp2).

model	precision	recall	f1	roc_auc	pr_auc	best_thresh old	thr_at_P $\geq$ 0 .90	recall_at_P $\geq$ 0 .90
LightGBM-Baseline	0.8737745 1	0.8638933 76	0.8688058 49	0.9876196 28	0.9325 1	0.79	0.87	0.838449111
Hybrid-AE+IF	0.8715334 42	0.8630856 22	0.8672889 61	0.9769623 45	0.8866 54	0.6	0.65	0.838449111
Hybrid-Contrastive+IF	0.8680387 41	0.8687399 03	0.8683891 8	0.9775895 87	0.8953 75	0.6	0.67	0.83723748

Source: Authors, (2026).

The results of Table 3 indicate the performance of anomaly-driven approaches to SSL that include an Autoencoder (AE) with an Isolation Forest (IF), and are not supervised integrated. Even though these techniques are effective in providing the detection of abnormal transaction patterns, they are less precise, and PR-AUC in comparison to the supervised baselines. It is shown that anomaly, and only knowledge fails to help differentiate between the fraud and the ordinary operation anomalies and hence the motivation towards hybrid knowledge integration and not solely depending on the use of the SSL.

IV.3 SSL FINE-TUNING AND STACKED KNOWLEDGE INTEGRATION (EXP3)

Table 4: Comparison between Baseline, SSL and Stacked Models (Best-F1 policy)

model	precision	recall	f1	roc_auc	pr_auc	best_threshold
LightGBM-Baseline	0.898103	0.822294	0.858528	0.988935	0.925073	0.89
SimCLR+Focal (SSL finetuned)	0.86358	0.851373	0.857433	0.986442	0.916541	0.46
Stacked (Baseline + SSL)	0.865478	0.862682	0.864078	0.987803	0.923521	0.95

Source: Authors, (2026).

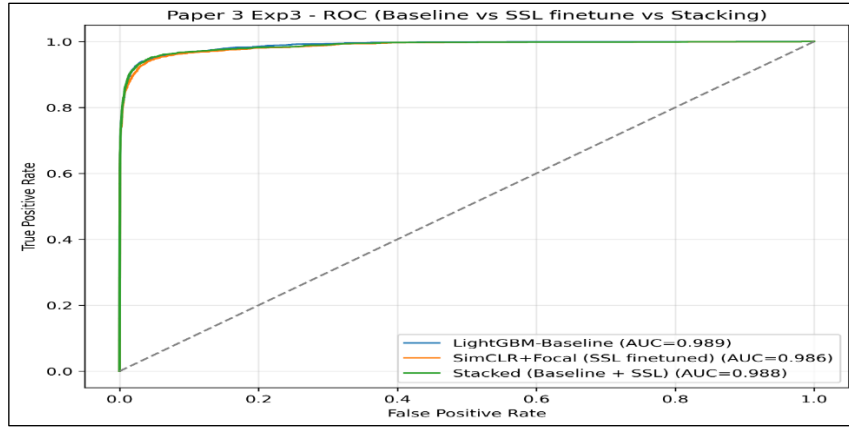


Figure 2: ROC Curves (Exp3).

Source: Authors, (2026).

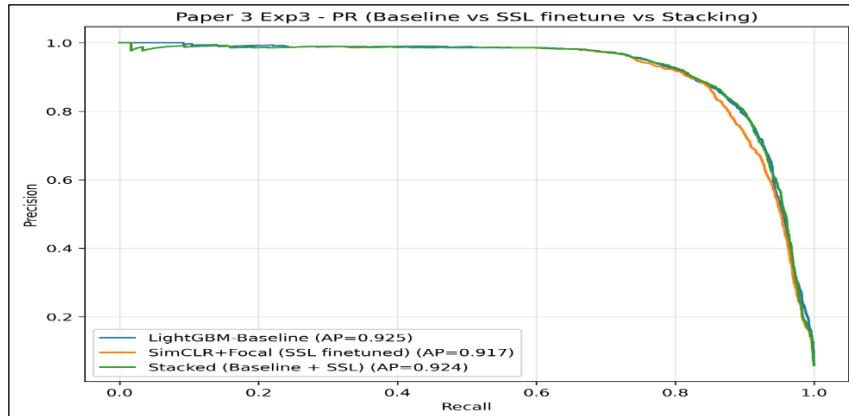


Figure 3: Precision-Recall Curves (Exp3).

Source: Authors, (2026).

As shown in Table 4 this experiment compares task aligned SSL, namely SimCLR with focal loss and a stacking approach, which combines supervised predictions with outputs of SSL. Figure 2 and 3 show that the stacked model generates a slight (around 45% to 55%) improvement in fraud recall compared with the baseline with a small F1 improvement (around 0.6). The ROC and PR curves demonstrate that stacking does not eliminate the discriminative sharpness of supervised learning and can extrapolate to a greater range of rare cases of frauds. These observations confirm that when used as representational support, one can add value to the use of SSL, but it does not replace the supervised core.

IV.4 SUPCON SSL ABLATION RESULTS (EXP4)

Table 5 SupCon SSL Ablation Results (Exp4).

model	precision	recall	f1	roc_auc	pr_auc	best_threshold
LightGBM-Baseline	0.896	0.814216	0.853153	0.989504	0.918041	0.91
SupCon+Focal (SSL finetuned)	0.90592	0.797254	0.84812	0.985089	0.910757	0.48
Stacked (Baseline + SupCon-SSL)	0.888792	0.823102	0.854687	0.986496	0.916789	0.97

Source: Authors, (2026).

As evidenced by Table 5, a stricter learning regimen of contrastive learning does not provide any extra benefits, and in fact, slightly decreases F1 and PR-AUC in some measurements. The adverse finding is important, which highlights that increased representational complexity does not always lead to enhanced knowledge of fraud, especially in the sparsely relational situations.

IV.5 OPERATION POLICY EVALUATION: PRECISION 6= 0.90 (EXP5)

Table 6 Precision- Constrained Evaluation (Exp5).

model	policy	threshold	precision	recall	f1	roc_auc	pr_auc
LightGBM-Baseline	best_f1	0.91	0.896	0.814216	0.853153	0.989504	0.918041
LightGBM-Baseline	precision $\geq$ 0.90	0.92	0.90144	0.808966	0.852703	0.989504	0.918041
SSL-Finetuned	best f1	0.53	0.907254	0.798061	0.849162	0.986619	0.910215
SSL-Finetuned	precision $\geq$ 0.90	0.54	0.912617	0.788772	0.846187	0.986619	0.910215
Stacked	best f1	0.97	0.898706	0.813409	0.853933	0.987591	0.916682
Stacked	precision $\geq$ 0.90	0.97	0.898706	0.813409	0.853933	0.987591	0.916682

Source: Authors, (2026).

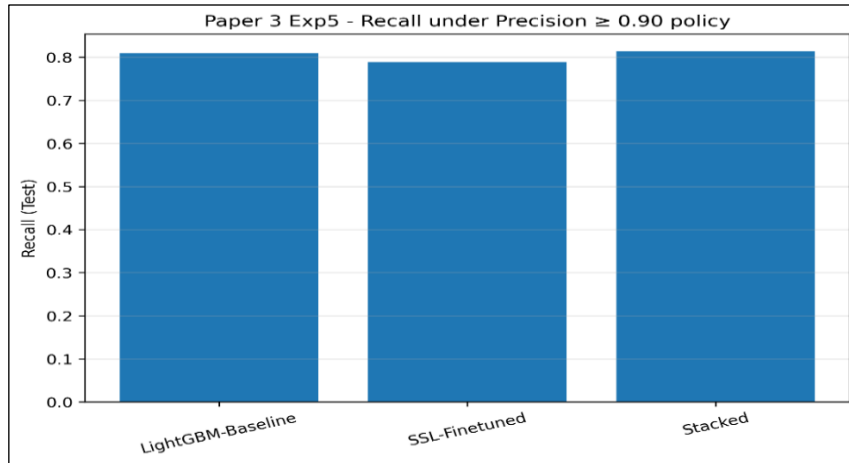


Figure 4. Recall with Precision 0.90 and above.

Source: Authors, (2026).

As shown in Table 6, in the setup of high precision penalties, all the SSL-based techniques outperform the supervised baseline. Both recall as well as F1 undergo slight decreases which depicts that the original model is already working near the optimal decision frontier. This observation brings out a significant operational conclusion such that the constraints on decisions may overshadow marginal gains in the algorithm, and knowledge saturation places constraints on the improvements that can be made by tuning to the threshold.

IV.6 SUMMARY OF EVIDENCE

The obtained results shown on tables and figures taken together demonstrate that although supervised models are able to achieve most of the knowledge that is of interest to the fraud problem, self-supervised learning only improves the recall when wisely combined and does not increase the overall performance in a consistent way. The evidence provides a subtle, empirical based insight into the timing of representation learning in providing decision value in large scale in fraud knowledge management system.

IV.7 DISCUSSION

IV.7.1 Why Self-Supervised Learning Saturates in Fraud Knowledge Extraction

It has been discovered that self-supervised learning does not invariably enhance fraud detection performance as compared to the robust supervised baselines, especially where there are serious operational restraint conditions. In terms of knowledge-management, this saturation is a representation of inherent characteristics of the issue of fraud detection and not the methodological deficiencies. Knowledge of frauds is frequently transaction-local and attribute-based in large scale payment systems, i.e., most discriminative information has already been included in structured characteristics like amount, timing, merchant category and payment channel. High signal Gradient-boosting models are good at leveraging these high signal attributes, and there is little residual information that the representation learning can rediscover. The main reason that regularities of usual behavior are captured using self-supervised learning is that the resulting representations are very useful in detecting an anomaly compared to the common trends. Nevertheless, numerous operational errors (e.g., technical errors, lack of balance, bad PIN attempts) often seem to be anomalous and at the same time do not constitute fraud.

Subsequently, anomaly-based representations bring ambiguity in their pure form to misalign regular process friction with criminal intent. This is why precision of standalone solutions based on the use of the SSL method remains reduced and why task alignment in knowledge representation is vital. The gains are also concentrated on recall and not on global discrimination metrics even when self-supervised embeddings are fine-tuned using labels of fraud and stacked together. This implies that there is no evidence that the marginal knowledge contributed to the boundary of rare fraud with by the use of the SSL, which enhances covering of the subtle or emerging cases, but does not transform the overall decision frontier. With a policy which is constrained to precision, the base model already runs in the Pareto-optimal region, with little more to be gained by representational refinements. Notably, this limitation is not overcome by heavier contrastive objectives and supervised contrastive learning, which implies that an increase in representational complexity does not ensure richer knowledge of fraud in the case of weak relations density and coordinated fraud structures.

Taken together, these results point to a key lesson of information and knowledge management, which is that representation learning can only be useful in situations where latent, underexploited structure is present in the underlying data. Knowledge saturation may happen in mature transactional data where the signals are strong at the attribute level, and model sophistication is less important to decision quality than are operational constraints.

IV.7.2 Managerial and Knowledge Implications

The results of this paper present a number of relevant ideas to managers, risk officers, and financial information and knowledge management systems designers. To begin with, the outcomes show that already when using powerful supervised learning models, a significant part of the operational knowledge related to fraud can be encoded as a result of structured transactional data. To managers, this is to say that there should be no automatic replacement of the existing supervised representations learning with new ones, but they must be assessed as to the extra knowledge they bring with particular decision goals like recall enhancement or span of early warning. Second, the small returns that have been realized by isolated self-managed anomaly detection point to a significant difference between operational anomalies and fraudulent behavior. This implies that the anomaly scores should be rather considered as contextual knowledge rather than direct fraud indicator when viewed through the prism of management. By adding these signals in decision-support systems, it is possible to prioritize these investigations, minimize analyst fatigue, and distinctions between system friction and real risk, increasing the efficiency of operations without increasing false-positive rates. Third, the recall gains obtained via stacked integration of supervised as well as self-supervised models support the usefulness of knowledge-fusion strategies.

Instead of leveraging models that keep growing in complexity, managers can enjoy the benefits of architectures that unite supplementary knowledge resources, i.e., rule based knowledge, supervised risk scores and self-governing behavioral representations, into core curation pipelines. The methodology helps in progressive evolution of the system and maintain transparency and regulatory adherence. Based on the knowledge-management perspective, the findings highlight the significance of alignment of the representation. In the cases where its representations are clearly aligned with downstream decision tasks and organizational objectives, self-supervised learning has the most significant impact. The burdened or more complicated representations do not necessarily provide better knowledge about decisions and can present interpretability problems. Lastly, the saturation in the situation of stringent accuracy requirements supports the importance of policy-based decision thresholds as the main artefacts of knowledge. Fraud management is effective, however, not just a matter of the accuracy of the models, but the operationalization of learned knowledge in governance structures, risk attitudes, and human-in-the-loop processes.

IV.7.3 Key Insights

- Fraud detection is not only a modelling problem, but also a knowledge problem.
- Most of the fraud-relevant knowledge in structured transaction systems in large-scale payment systems are already embedded in robust supervised models.
- Self-managed learning is a value addition that is selective.
- Fraud discrimination cannot be done by stand-alone anomaly detection, however with supervised models task-aligned self-supervised representations can enhance recall.
- Model replacement is of inferior quality to knowledge integration, and by combining supervised and self-supervised output is more dependable on gains than using self-supervised models alone.
- The operational constraints restrict the gains of algorithms, with very high accuracy conditions (e.g., in excess of 0.90) no further improvements are obtained with increased representational complexity, indicating there is a saturation point in the extraction of fraud knowledge.
- Managerial relevance is a factor of alignment.
- The knowledge management of effective fraud and fraud prevention should be aligned with the decision policies, governance requirements and human in the loop procedures instead of undertaking complexity itself.

V. CONCLUSION

The present investigation has taken the task of fraud detection not as a local competition between competing algorithms, but as a complex knowledge-representation and decision-making problem. We used a massive, real-world dataset of payment volumes that is defined by deep class imbalance and an ecology of operation indicators to perform a methodological appraisal of there being supervised learning, self-bravo anomaly recognition, and mottled knowledge-combination techniques. We find empirically that the existing supervised models already enshrine a considerable share of fraud-relevant knowledge based on structured transactional features, and have nearly optimal discrimination performance. On the other hand, self-managed learning, when applied alone, does not always separate fraud and legitimate operational anomalies but it can introduce uncertainty into subsequent decision making. The key outcome of the research is the evidence that self-managed representations only add value when they are sensitive to the ultimate decision goals, and they are used as supplementary sources of knowledge. Task-aligned self-supervised learning can be used to improve the fraud recall of rare cases with stacked representations of such representations with supervised models without adversely impacting operational reliability. However, such benefits are automatically limited by the nature of the data. Even with the high level of precision required in the real world of fraud, representational complexity does not help any further, which highlights a practice limit to algorithmic returns as soon as the knowledge boundary is over. In terms of information and knowledge management, these findings send an essential signal to researchers and practitioners the fact that the complexity of any representation is not an assurance of better quality knowledge of decisions. Strong fraud systems have to achieve a balance between representational richness, interpretability and policy based restraint of decision outcomes. This work has its benefit in that it establishes the exact circumstances in which self-supervised learning can be used to supplement fraud knowledge and vice versa.

We provide a template balancing theory and practice by giving an evidence-based model that can be used to design resilient, scalable and governable fraud knowledge-management systems. In laying an accentuation on integration, alignment, as well as decision relevance, the research advances the academic discussion of how emergent learning paradigms ought to be implemented in modern financial information infrastructures. It promotes a shift of the parochial algorithmic competitions into knowledge-based holistic approaches, sensitive to organisational goals, regulatory provisions and the demands of real-time operational decision making.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Boumedyen Shannaq.

Methodology: Boumedyen Shannaq, Basel Bani-Ismail.

Investigation: Boumedyen Shannaq, Nour Eldin Elshaiekh.

Discussion of results: Boumedyen Shannaq.

Writing – Original Draft: Boumedyen Shannaq.

Writing – Review and Editing: Boumedyen Shannaq, Marwan Alshar'e, Oualid Ali.

Resources: Boumedyen Shannaq, Nour Eldin Elshaiekh, Basel Bani-Ismail, Marwan Alshar'e, Oualid Ali.

Supervision: Boumedyen Shannaq.

Approval of the final text: Boumedyen Shannaq, Nour Eldin Elshaiekh, Basel Bani-Ismail, Marwan Alshar'e, Oualid Ali.

VII. REFERENCES

- [1] H. M. Farkhan and L. Indahdewi, "The Implementation of Knowledge Management on Fraud Prevention of PT Bahagia Idkho Mandiri," *devotion*, vol. 4, no. 6, pp. 1270–1281, May 2023, doi: 10.59188/devotion.v4i6.482.
- [2] W.-Y. Wang, W.-W. Du, D. Xu, W. Wang, and W.-C. Peng, "A survey on self-supervised learning for non-sequential tabular data," *Mach Learn*, vol. 114, no. 1, p. 16, Jan. 2025, doi: 10.1007/s10994-024-06674-0.
- [3] K. M. Sujon, R. Hassan, K. Choi, and M. A. Samad, "Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models," *J Big Data*, vol. 12, no. 1, p. 268, Dec. 2025, doi: 10.1186/s40537-025-01313-4.
- [4] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable Machine Learning in Credit Risk Management," *Comput Econ*, vol. 57, no. 1, pp. 203–216, Jan. 2021, doi: 10.1007/s10614-020-10042-0.
- [5] B. Shannaq, "A Novel Algorithm for Differentiating Authorized Users from Fraudsters via Mobile Keypad Input Patterns During Password Updates," *TEM Journal*, pp. 768–778, Feb. 2025, doi: 10.18421/TEM141-68.
- [6] B. Shannaq, "Advancing User Authentication in Cloud Computing: Automated Smart Application as a Multi-Factor Approach Beyond Extended-Levenshtein Distance," *TEM Journal*, pp. 1140–1151, May 2025, doi: 10.18421/TEM142-15.
- [7] B. Shannaq, M. Shakir, and A. Al-Zeidi, "Predicting User Originality in Password Activities Using Machine Learning," in *Tech Fusion in Business and Society*, vol. 234, R. K. Hamdan, Ed., in *Studies in Systems, Decision and Control*, vol. 234, Cham: Springer Nature Switzerland, 2025, pp. 59–69, doi: 10.1007/978-3-031-84636-6_6.
- [8] S. Li, J. Yu, Y. Lu, G. Yang, X. Du, and S. Liu, "Self-supervised enhanced denoising diffusion for anomaly detection," *Information Sciences*, vol. 669, p. 120612, May 2024, doi: 10.1016/j.ins.2024.120612.
- [9] T. Sattarov, M. Schreyer, and D. Borth, "Diffusion-Scheduled Denoising Autoencoders for Anomaly Detection in Tabular Data," 2025, arXiv, doi: 10.48550/ARXIV.2508.00758.
- [10] G. Cignoni, A. Cossu, A. Gomez-Villa, J. van de Weijer, and A. Carta, "CLA: Latent Alignment for Online Continual Self-Supervised Learning," 2025, arXiv, doi: 10.48550/ARXIV.2507.10434.
- [11] S. Saeed, S. A. Suayyid, M. S. Al-Ghamdi, H. Al-Muhaisen, and A. M. Almuhaideb, "A Systematic Literature Review on Cyber Threat Intelligence for Organizational Cybersecurity Resilience," *Sensors*, vol. 23, no. 16, p. 7273, 2023, doi: 10.3390/s23167273.
- [12] B. Shannaq and M. Shakir, "Enhancing Security with Multi-Factor User Behavior Identification via Longest Common Subsequence Analysis," *Informatica*, vol. 48, no. 16, pp. 73–82, 2024, doi: https://doi.org/10.31449/inf.v48i19.6529.
- [13] C. Huang, L. Xia, X. Wang, X. He, and D. Yin, "Self-Supervised Learning for Recommendation," in *Proceedings of the ACM Web Conference 2022 (WWW '22)*, New York, NY, USA: Association for Computing Machinery (ACM), 2022, pp. 1847–1858, doi: 10.1145/3511808.3557506.
- [14] L. Diana and P. Dini, "Overview on Intrusion Detection Systems for Computers Networking Security," *Computers*, vol. 14, no. 3, p. 87, 2025, doi: 10.3390/computers14030087.
- [15] C. Bekkaye, H. Oukhouya, T. Zari, R. Guerbaz, and H. El Bouanani, "Generative hybrid models for fraud detection in auto insurance with a comparative analysis of VAE, GAN, and diffusion approaches," *Discov Artif Intell*, vol. 5, no. 1, p. 313, Nov. 2025, doi: 10.1007/s44163-025-00574-5.
- [16] M. Chabbouh, S. Bechikh, E. Mezura-Montes, and L. Ben Said, "curve for classifying imbalanced multi-class data," *J Heuristics*, vol. 31, no. 1, p. 9, Mar. 2025, doi: 10.1007/s10732-024-09544-z.
- [17] L. Hernandez Aros, L. X. Bustamante Molano, F. Gutierrez-Portela, J. J. Moreno Hernandez, and M. S. Rodríguez Barrero, "Financial fraud detection through the application of machine learning techniques: a literature review," *Humanit Soc Sci Commun*, vol. 11, no. 1, p. 1130, Sep. 2024, doi: 10.1057/s41599-024-03606-0.
- [18] I. R. Al-Shamsi, B. Shannaq, R. Adebaiye, and T. Owusu, "Exploring biometric attendance technology in the Arab academic environment: Insights into faculty loyalty and educational performance in policy initiatives," *J. Infr. Policy. Dev.*, vol. 8, no. 9, p. 6991, Sep. 2024, doi: 10.24294/jipd.v8i9.6991.

- [19] A. A. Compagnino et al., "An Introduction to Machine Learning Methods for Fraud Detection," *Applied Sciences*, vol. 15, no. 21, p. 11787, Nov. 2025, doi: 10.3390/app152111787.
- [20] H. Yang, Z. Shukur, and S. Sahran, "A Review of Artificial Intelligence for Financial Fraud Detection," *Applied Sciences*, vol. 16, no. 4, p. 1931, 2026, doi: 10.3390/app16041931.
- [21] T. Miftahshudur, H. M. Sahin, B. Grieve, and H. Yin, "A Survey of Methods for Addressing Imbalance Data Problems in Agriculture Applications," *Remote Sensing*, vol. 17, no. 3, p. 454, Jan. 2025, doi: 10.3390/rs17030454.
- [22] S. Visbeek, E. Acar, and F. den Hengst, "Explainable Fraud Detection with Deep Symbolic Classification," Dec. 01, 2023, arXiv: arXiv:2312.00586. doi: 10.48550/arXiv.2312.00586.
- [23] C. van der Cruysen and M. Plooi, "Changing Payment Patterns at Point-of-Sale: Their Drivers," *De Nederlandsche Bank Working Paper*, no. 471, Apr. 2015, doi: 10.2139/ssrn.2602219.
- [24] B. Shannaq and O. Ali, "Credit Card Fraud Detection: An Explainable AI Approach for Securing Financial Transactions," in *2025 3rd International Conference on Business Analytics for Technology and Security (ICBATS)*, Dubai, United Arab Emirates: IEEE, May 2025, pp. 1–6. doi: 10.1109/ICBATS66542.2025.11258465.
- [25] S. B. Atitallah, C. B. Rabah, M. Driss, W. Boulila, and A. Koubaa, "Self-Supervised Learning for Graph-Structured Data in Healthcare Applications: A Comprehensive Review," 2024, arXiv. doi: 10.48550/ARXIV.2412.05312.
- [26] M. M. Abdulrazzaq et al., "Consequential Advancements of Self-Supervised Learning (SSL) in Deep Learning Contexts," *Mathematics*, vol. 12, no. 5, p. 758, Mar. 2024, doi: 10.3390/math12050758.
- [27] T. Richter, M. Bahrami, Y. Xia, D. S. Fischer, and F. J. Theis, "Delineating the effective use of self-supervised learning in single-cell genomics," *Nat Mach Intell*, vol. 7, no. 1, pp. 68–78, Dec. 2024, doi: 10.1038/s42256-024-00934-3.
- [28] B. Shannaq, M. A. Talab, M. Shakir, M. T. Sheker, and A. M. Farhan, "Machine learning model for managing the insider attacks in big data," presented at the *THE SECOND INTERNATIONAL CONFERENCE ON EMERGING TECHNOLOGY TRENDS IN INTERNET OF THINGS AND COMPUTING*, Ramadi, Iraq, 2023, p. 020013. doi: 10.1063/5.0188358.
- [29] Z. Wang et al., "A Comprehensive Survey on Data Augmentation," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 38, no. 1, pp. 47–66, Jan. 2026, doi: 10.1109/TKDE.2025.3622600
- [30] D. Vijayanand and G. S. Smrithy, "Explainable AI - enhanced ensemble learning for financial fraud detection in mobile money transactions," *Intelligent Decision Technologies*, vol. 19, no. 1, pp. 52–67, Jan. 2025, doi: 10.1177/18724981241289751.
- [31] F. G. Popa and V. Muresan, "Artificial Intelligence in Finance: From Market Prediction to Macroeconomic and Firm-Level Forecasting," *AI*, vol. 6, no. 11, p. 295, Nov. 2025, doi: 10.3390/ai6110295.
- [32] B. Shannaq and O. Ali, "Leveraging Machine Learning for Predictive Sustainability in Business Operations: A Classification Approach to Optimize Sustainable Resource Management," in *Sustainable Data Management*, R. K. Hamdan, Ed., *Studies in Big Data*, vol. 171, Cham, Switzerland: Springer, 2025, doi: 10.1007/978-3-031-83911-5_57.
- [33] D. D. Pezze, E. Anello, C. Masiero, and G. A. Susto, "Continual learning approaches for anomaly detection," *Evolving Systems*, vol. 16, no. 4, p. 111, Dec. 2025, doi: 10.1007/s12530-025-09732-7.
- [34] X. Chen, Z. Liu, S. Xie, and K. He, "Deconstructing Denoising Diffusion Models for Self-Supervised Learning," 2024, arXiv. doi: 10.48550/ARXIV.2401.14404.
- [35] S. Khodabandehlou and A. H. Golpayegani, "FiFrauD: Unsupervised Financial Fraud Detection in Dynamic Graph Streams," *ACM Trans. Knowl. Discov. Data*, vol. 18, no. 5, pp. 1–29, Jun. 2024, doi: 10.1145/3641857.
- [36] D. Gupta, P. Chugh, K. Sood, and S. Grima, "Explainable Artificial Intelligence in Financial Fraud Detection: A Systematic Review and FinTech-Oriented ADO–TCCM Meta-Framework for Trust, Governance, and Transparency," *FinTech*, vol. 5, no. 3, p. 60, 2026, doi: 10.3390/fintech5030060.
- [37] Z. Xiao, "Research on financial fraud detection method based on graph neural network," in *International Conference on Algorithms, Image Processing, and Deep Learning (AIPDL 2025)*, X. Kong and Y. Y. Tang, Eds., Macau, China: SPIE, Aug. 2025, p. 8. doi: 10.1117/12.3078631.
- [38] Q. Sha, T. Tang, X. Du, J. Liu, Y. Wang, and Y. Sheng, "Detecting Credit Card Fraud via Heterogeneous Graph Neural Networks with Graph Attention," 2025, arXiv. doi: 10.48550/ARXIV.2504.08183.
- [39] B. Shannaq, D. K. Muniyanayaka, O. Ali, B. Bani-Ismail, and S. A. Maqbali, "Exploring the role of machine learning models in risk assessment models for developed organizations' management decision policies," *J. Infrastruct. Policy Dev.*, vol. 8, no. 13, p. 9364, 2024, doi: 10.24294/jipd9364.
- [40] I. D. Mienye and T. G. Swart, "A Comprehensive Review of Deep Learning: Architectures, Recent Advances, and Applications," *Information*, vol. 15, no. 12, p. 755, 2024, doi: 10.3390/info15120755.
- [41] Y. Hu, K. Xiao, L. Luo, and L. Chen, "An XGBoost-Based Intrusion Detection Framework with Interpretability Analysis for IoT Networks," *Applied Sciences*, vol. 16, no. 2, p. 980, Jan. 2026, doi: 10.3390/app16020980.
- [42] M. Imani, M. Joudaki, A. Bagheri, and H. R. Arabia, "Why ROC-AUC Is Misleading for Highly Imbalanced Data: In-Depth Evaluation of MCC, F2-Score, H-Measure, and AUC-Based Metrics Across Diverse Classifiers," *Technologies*, vol. 14, no. 1, p. 54, Jan. 2026, doi: 10.3390/technologies14010054.
- [43] T. Saito and M. Rehmsmeier, "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PLoS ONE*, vol. 10, no. 3, p. e0118432, Mar. 2015, doi: 10.1371/journal.pone.0118432.
- [44] "Financial Transactions Dataset: Analytics." Accessed: Feb. 02, 2026. [Online]. Available: <https://www.kaggle.com/datasets/computingvictor/transactions-fraud-datasets>