

الجمهورية الشعبية الديمقراطية الجزائرية
People's Democratic Republic of Algeria
وزارة التعليم العالي و البحث العلمي
Ministry of Higher Education and Scientific Research
المدرسة العليا للإعلام الآلي 8 ماي 1945 - سيدي بلعباس
Higher School of Computer Science
8 Mai 1945 - Sidi Bel Abbès



Master's Thesis

To obtain the diploma of Master's Degree

Field of Study: Computer Science

Specialization: Information Systems and Web (SIW)

Theme

**Fraud Detection in Online Payment Systems Based on
Artificial Intelligence and Machine Learning Techniques**

Presented by
Djemai Oumima
Taibi Narimane

Defended on: **June, 2026**
In front of the jury composed of

Dr. ALLAL Lamia.
Dr. BEDJAOUI Mohamed
Dr. CHIKH Asma
Dr. KECHAR Mohamed

President of the Jury
Thesis Supervisor
Examiner
Representative of the incubator,

Academic Year: 2025/2026

Abstract

Online payment fraud poses a major threat to banks, businesses, and customers worldwide, resulting in significant financial losses each year. Traditional rule-based detection systems are increasingly ineffective against sophisticated and evolving fraud schemes.

This master's thesis presents a state-of-the-art review of Artificial Intelligence (AI) and Machine Learning (ML) techniques for online payment and credit card fraud detection. Recent studies published between 2020 and 2026 were analyzed across four categories: Machine Learning, Ensemble Learning, Deep Learning, Anomaly Detection with Risk Management, and Systematic Literature Reviews. For each study, detection methods, datasets, preprocessing techniques, evaluation metrics, strengths, and limitations were examined.

The review shows that Random Forest remains a highly effective ML algorithm, particularly when combined with feature selection and optimization techniques. Deep Learning models such as LSTM and GRU perform well on sequential transaction data, while ensemble methods consistently outperform individual classifiers. Studies also demonstrate that combining anomaly detection with risk modelling can significantly reduce fraud losses while maintaining low false positive rates. Furthermore, the literature indicates a strong focus on supervised learning, with limited research on semi-supervised, reinforcement learning, and explainable AI approaches.

Several challenges persist, including extreme class imbalance, over-reliance on accuracy as an evaluation metric, computational constraints for real-time deployment, and limited model interpretability required for regulatory compliance.

The review concludes that no single technique can address all fraud detection challenges. Hybrid approaches that combine complementary methods appear to offer the greatest potential for improving detection performance and practical deployment. These findings provide a foundation for future research and the development of intelligent fraud detection systems.

Keywords: online payment fraud, credit card fraud, machine learning, deep learning, anomaly detection, ensemble learning, systematic review.

Résumé

La fraude dans les paiements en ligne est devenue une menace sérieuse pour les banques, les commerçants et les clients du monde entier, causant des milliards de dollars de pertes annuelles. Les méthodes traditionnelles basées sur des règles fixes ne sont plus suffisantes face à des stratégies de fraude de plus en plus sophistiquées.

Ce mémoire propose une revue complète de l'état de l'art des techniques d'intelligence artificielle et d'apprentissage automatique appliquées à la détection de fraude dans les paiements en ligne et par carte de crédit. Une sélection d'études récentes publiées entre 2020 et 2026 a été analysée, couvrant quatre catégories principales : les approches basées sur l'apprentissage automatique, l'apprentissage ensembliste, l'apprentissage profond, la détection d'anomalies combinée à la gestion des risques, et les revues systématiques.

L'analyse révèle que Random Forest reste l'un des algorithmes les plus robustes, atteignant des taux de précision élevés. Combiné avec des techniques d'optimisation comme les algorithmes génétiques, ses performances s'améliorent davantage. LSTM montre de bonnes performances sur les données séquentielles, tandis que GRU offre un bon équilibre entre performance et complexité. Les méthodes ensemblistes surpassent constamment les classificateurs uniques.

L'intégration de la détection d'anomalies avec un modèle de triage économique et un modèle de risque a réduit considérablement les pertes tout en maintenant de faibles taux de faux positifs. Les revues systématiques indiquent que l'apprentissage supervisé domine la littérature, tandis que l'apprentissage semi-supervisé et par renforcement restent largement inexplorés. Les algorithmes d'optimisation méta-heuristique ont atteint les niveaux de précision les plus élevés.

Cependant, plusieurs défis persistent : un déséquilibre extrême des classes (moins d'un pour cent de transactions frauduleuses), une dépendance excessive à la précision comme seule métrique, une complexité computationnelle élevée pour le déploiement en temps réel, et une absence de méthodes d'IA explicable, essentielles pour la conformité réglementaire dans le secteur bancaire.

La revue conclut qu'aucune technique unique ne peut répondre à tous les défis. Les approches hybrides combinant des techniques complémentaires semblent être la direction la plus prometteuse. Ces résultats fournissent une base solide pour la conception de systèmes de détection de fraude intelligents.

Mots-clés: fraude paiement en ligne, fraude carte de crédit, apprentissage automatique, apprentissage profond, détection d'anomalies, apprentissage ensembliste, revue systématique.

الملخص

أصبح الاحتيال في المدفوعات عبر الإنترنت تهديداً خطيراً للبنوك والتجار والعملاء في جميع أنحاء العالم، مما يتسبب في خسائر تبلغ مليارات الدولارات سنوياً. لم تعد طرق كشف الاحتيال التقليدية القائمة على قواعد ثابتة كافية لمواجهة استراتيجيات الاحتيال المتطورة والمتغيرة باستمرار.

تقدم هذه المذكرة مراجعة شاملة لحالة الفن لتقنيات الذكاء الاصطناعي والتعلم الآلي المطبقة على كشف الاحتيال في المدفوعات الإلكترونية وبطاقات الائتمان. تم تحليل مجموعة من الدراسات الحديثة المنشورة بين 2020 و 2026. تغطي هذه الدراسات أربع فئات رئيسية: المناهج القائمة على التعلم الآلي، والتعلم التجميعي، والتعلم العميق، وكشف الحالات الشاذة مع إدارة المخاطر، والمراجعات المنهجية. لكل دراسة، تم فحص خوارزميات الكشف، ومجموعات البيانات، وتقنيات المعالجة المسبقة، وطرق اختيار الخصائص، ومقاييس التقييم، ونقاط القوة والضعف.

يكشف التحليل عدة نتائج رئيسية. Random Forest لا تزال واحدة من أقوى خوارزميات التعلم الآلي، حيث تحقق معدلات دقة عالية في مهام كشف الاحتيال. عند دمجها مع تقنيات التحسين مثل الخوارزميات الجينية لاختيار الخصائص، تتحسن الأداء بشكل أكبر. من بين بنى التعلم العميق، يُظهر LSTM أداءً قوياً على بيانات المعاملات المتسلسلة، بينما يقدم GRU توازناً جيداً بين الأداء والتعقيد. الأساليب التجميعية، بما في ذلك Bagging و Boosting و Stacking، تتفوق باستمرار على المصنفات الفردية.

أظهر دمج كشف الحالات الشاذة مع نموذج فرز اقتصادي ونموذج مخاطر نتائج مهمة، حيث قلل خسائر الاحتيال بشكل كبير مع الحفاظ على معدلات منخفضة من الإيجابيات الكاذبة. تشير المراجعات المنهجية في هذا المجال إلى أن التعلم الخاضع للإشراف يهيمن على الأدبيات، بينما لا يزال التعلم شبه الخاضع للإشراف والتعلم التعزيزي غير مستكشفين إلى حد كبير. حققت خوارزميات التحسين الميتاهيورستيك بعضاً من أعلى مستويات الدقة المسجلة في كشف الاحتيال.

ومع ذلك، تم تحديد العديد من التحديات المستمرة في الأدبيات. وتشمل هذه عدم التوازن الشديد في الفئات، حيث تمثل المعاملات الاحتيالية عادة أقل من واحد في المائة من جميع المعاملات؛ والاعتماد المفرط على الدقة كقياس تقييم وحيد، وهو أمر إشكالي للبيانات غير المتوازنة؛ والتعقيد الحسابي العالي للنشر في الوقت الفعلي؛ والغياب العام لأساليب الذكاء الاصطناعي القابل للتفسير، وهي أمور ضرورية للامتثال التنظيمي في القطاع المصرفي.

تخلص المراجعة إلى أنه لا توجد تقنية واحدة يمكنها معالجة جميع تحديات كشف الاحتيال بشكل فعال. تبدو المناهج الهجينة التي تجمع بين تقنيات متكاملة الاتجاه الأكثر وعداً للبحث المستقبلي. توفر هذه النتائج أساساً متيناً لتصميم وتنفيذ أنظمة كشف احتيال ذكية.

الكلمات المفتاحية: احتيال المدفوعات الإلكترونية، احتيال بطاقات الائتمان، تعلم آلي، تعلم عميق، كشف الحالات الشاذة، تعلم تجميعي، مراجعة منهجية