# *AVIS DE SOUTENANCE DE THESE*

*Le Doyen de la Faculté des Sciences Dhar El Mahraz –Fès – annonce que*

Mme (elle) **EL HLOULI Fatima Zohra**
Soutiendra : **le Vendredi 23/02/2024** à **15H00**
*Lieu :* ***FSDM – Centre Visioconférence***

*Une thèse intitulée :*

## Fraud Detection Using Deep Learning Architectures

*En vue d'obtenir le* ***Doctorat***

*FD :* ***Sciences et Technologies de l'Information et de la Communication***
*Spécialité :* ***Informatique***

***Devant le jury composé comme suit :***

| Nom et prénom | Etablissement | Grade | Qualité |
|---|---|---|---|
| Pr TAIRI Hamid | Faculté des Sciences Dhar El Mahraz, Fès | PES | Président |
| Pr AGHOUTANE Badraddine | Faculté des Sciences, Meknès | PES | Rapporteur & Examinateur |
| Pr KHAMJANE Aziz | Ecole Nationale des Sciences Appliquées, Al Hoceima | PH | Rapporteur & Examinateur |
| Pr NFAOUI El Habib | Faculté des Sciences Dhar El Mahraz, Fès | PES | Rapporteur & Examinateur |
| Pr EL FAZAZY Khalid | Faculté des Sciences Dhar El Mahraz, Fès | PH | Examinateur |
| Pr MAHRAZ Mohamed Adnane | Faculté des Sciences Dhar El Mahraz, Fès | PH | Examinateur |
| Pr YAHYAOUY Ali | Faculté des Sciences Dhar El Mahraz, Fès | PES | Examinateur |
| Pr RIFFI Jamal | Faculté des Sciences Dhar El Mahraz, Fès | PH | Directeur de thèse |

## Résumé

La pandémie de coronavirus a engendré une croissance significative des transactions bancaires effectuées en ligne. Cette évolution s'explique par les restrictions de déplacement et les mesures de distanciation sociale imposées pendant la pandémie, poussant de nombreuses personnes à privilégier les paiements en ligne et les transactions à distance pour leurs besoins quotidiens. La concentration accrue sur la détection et la prévention de la fraude, notamment dans le domaine des émetteurs de cartes de crédit, découle des importantes pertes financières annuelles dues à une utilisation frauduleuse des cartes. La réduction de la fraude peut engendrer d'importantes économies sur les coûts opérationnels, ce qui suscite des recherches approfondies dans ce domaine.

Dans la détection de fraudes, l'objectif est d'obtenir un modèle performant qui identifie avec précision les transactions frauduleuses tout en minimisant les faux positifs. Cependant, ce défi implique également de considérer la complexité du modèle et le temps nécessaire à la détection. Un modèle performant doit trouver un équilibre entre ces trois aspects. Un modèle trop complexe peut être difficile à interpréter, tandis qu'un modèle trop simple peut compromettre la précision. De plus, le temps nécessaire pour détecter une fraude est crucial, surtout dans des situations nécessitant des réponses rapides. Ainsi, le défi consiste à trouver un modèle efficace, suffisamment précis et rapide, adapté aux besoins spécifiques de détection de fraude, tout en tenant compte des contraintes de ressources et des impératifs de réactivité.

Les réseaux de neurones artificiels se sont révélés prometteurs pour la détection des fraudes. Ils fournissent une base solide pour créer des modèles prédictifs basés sur des données structurées, tandis que les méthodes d'apprentissage profond offrent des capacités avancées pour analyser des données non structurées ou séquentielles, découvrir des modèles cachés et détecter des stratagèmes de fraude sophistiqués.

Quant aux méthodes d'optimisation telles que le Dandelion, Particleswarm et le Group search algorithme, elles sont souvent utilisées pour rechercher efficacement des solutions dans des espaces de recherche complexes, pour trouver les paramètres optimaux dans les modèles d'apprentissage automatique ou pour ajuster les hyperparametres des algorithmes.

En utilisant ces méthodes, notre approche semble inclure une combinaison d'architectures et techniques d'apprentissage appliquées à des ensembles de données déséquilibrés et équilibrés qui sont fréquents dans les scénarios de détection de fraudes. Pour établir un cadre de comparaison, nous avons mené des recherches sur la sélection des caractéristiques, le prétraitement et les différentes mesures de performance.

Nos expériences ont utilisé des données réelles de transactions par carte de crédit et ont employé des techniques distinctes d'apprentissage automatique et profond tel que le réseau de neurones entièrement connecté (FCNN), ainsi que deux méthodes innovantes basé sur : l'Extreme Learning Machine (ELM) et l'Autoencodeur (AE).

**Mots-clés :** Fraude, Cartes de crédit, Prétraitement de données, Données déséquilibrées, Apprentissage supervisé, Réseaux neuronaux entièrement connecté, Extremelearning machine, Autoencoder, Majorityvoting, Dandelionalgorithm.

# *FRAUD DETECTION USING DEEP LEARNING ARCHITECTURES*

**Abstract :**

The coronavirus pandemic has led to significant growth in online banking transactions. The travel restrictions and social isolation measures put in place during the pandemic are what explain this development because they led many people to prefer online payments and remote transactions for their daily needs. The increased focus on fraud detection and prevention, particularly around credit card issuers, stems from the significant annual financial losses due to fraudulent card use. Reducing fraud can result in significant operational cost savings, prompting extensive research in this area.

In fraud detection, the goal is to obtain a high-performance model that accurately identifies fraudulent transactions while minimizing false positives. However, this challenge also involves considering the complexity of the model and the time required for detection. A successful model must find a balance between these three aspects. A model that is too complex can be difficult to interpret, while a model that is too simple can compromise accuracy. Additionally, the time it takes to detect fraud is crucial, especially in situations requiring rapid responses. Thus, the challenge consists of finding an effective model that is sufficiently precise and fast, adapted to the specific needs of fraud detection, while considering resource constraints and responsiveness requirements.

Artificial neural networks have shown promise for fraud detection. Machine learning and deep learning are two branches of artificial intelligence that have revolutionized the field of fraud detection. Deep learning, a subset of machine learning, involves neural networks with many layers. Both machine learning and deep learning play crucial roles in fraud detection systems. Machine learning techniques provide a solid foundation for building predictive models based on structured data, while deep learning methods offer advanced capabilities for analyzing unstructured or sequential data, uncovering hidden patterns, and detecting sophisticated fraud schemes.

Ensemble learning methods are approaches that combine predictions from multiple learning models to improve overall accuracy. These methods aim to reduce individual model errors by combining their predictions strategically, for example, by majority voting or averaging predictions.

As for optimization methods such as Dandelion, Particle swarm, and Group search algorithms, they are often used to efficiently search for solutions in complex search spaces. These techniques draw inspiration from the collective behavior of particles or groups in nature to explore and optimize sets of parameters, for example, to find optimal parameters in machine learning models or to tune hyperparameters of algorithms.

Using these methods, our approach appears to include a combination of architectures based on ensemble learning techniques, which aim to improve the robustness and performance of the models, as well as specific optimization methods to find configurations of models or parameters that are more efficient and accurate. To address this issue, we propose using dynamic machine-learning approaches to model the nature of card transaction data.

To establish a framework for comparison, we researched feature selection, preprocessing, and performance metrics. Imbalanced data occurs when the majority of transactions are legitimate (non-fraudulent), while instances of fraud make up a small fraction of all transactions. This disparity creates an imbalance in the data, which can pose challenges for machine learning models. Algorithms may tend to favor the majority class, thus leading to lower performance in detecting the minority class, such as fraud.

These efforts allow us to effectively compare results obtained from machine learning methods applied to unbalanced and balanced datasets, common in fraud detection scenarios. Our experiments used real credit card transaction data and employed distinct deep learning and machine learning techniques: a fully connected neural network (FCNN), as well as two innovative methods, the Extreme Learning Machine (ELM) and Autoencoder (AE). In the coming chapters, an in-depth analysis will be presented to explain the motivations behind the choice of these techniques. The specific reasons why these algorithms were selected will be discussed in detail, highlighting their abilities to deal with imbalanced data, model complex temporal sequences, or provide increased accuracy in fraud detection.

**Key Words :**
Fraud, Credit cards, Data preprocessing, Imbalanced data, Supervised learning, Fully Connected Neural networks, Extreme learning machine, Autoencoder, Majority voting, Dandelion algorithm.