

الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي - 08 ماي 1945 - بسيدي بلعباس
École Supérieure en Informatique, 08 Mai 1945- Sidi Bel Abbès



MEMOIRE

En Vue de l'obtention du diplôme de **Master**

Filière: **Informatique**

Spécialité: **Intelligence Artificielle et Sciences des Données (IASD)**

Thème

Robust Federated Learning under Statistical Heterogeneity
A Survey on Aggregation, Client Selection, Clustering and Security
Mechanisms

Présenté par: **Rahmoun Wassim**

Soutenu le: **06/07/2026**

Devant le jury composé de:

Someone

Someone

Dr. Dif Nassima

Dr. Ouchani Samir

Président

Examinateur

Encadreur

Co-Encadreur

Année Universitaire : 2025 / 2026

Abstract

The rapid growth of IoT devices has created a data environment that conventional, centralised ML is poorly equipped to handle: data that is distributed across billions of devices, sensitive in nature, and subject to privacy regulations that restrict how it can be collected and stored. FL addresses this by training a shared model across distributed clients without ever centralising their raw data, making it a natural fit for IoT deployments. This thesis is structured in two parts. The first establishes the conceptual and technical foundations needed to understand FL in this context, covering the IoT network architecture that motivates it, the ML and DL principles it builds on, the FL protocol and its main algorithmic variants, and the SI and evolutionary optimisation techniques increasingly used to address its combinatorial design problems. The second surveys the state of the art on robust FL across four central challenges: statistical heterogeneity, client selection, model clustering, and security threats. A comparative discussion brings these four areas together under a common set of evaluation criteria, revealing recurring structural gaps in the literature, most notably a persistent trade-off between robustness to heterogeneous data and resilience to malicious participants, an underuse of hierarchical architectures despite their demonstrated advantages, and a lack of frameworks that treat client selection, clustering, and aggregation as a jointly designed pipeline rather than independent components.

Keywords— Federated Learning, Hierarchical Architecture, Statistical Heterogeneity, Distributed Computing , Swarm Intelligence, Evolutionary Aggregation

Résumé

L'émergence rapide de IoT a engendré un environnement de données que le ML centralisé traditionnel est mal équipé pour prendre en charge : des données dispersées sur des milliards d'appareils, de nature sensible, et soumises à des réglementations sur la vie privée qui encadrent strictement leur collecte et leur stockage. Le FL répond à ce défi en entraînant un modèle partagé auprès de clients distribués sans jamais centraliser leurs données brutes, ce qui en fait une approche naturellement adaptée aux déploiements IoT. Ce mémoire s'articule en deux parties. La première pose les fondements conceptuels et techniques nécessaires à la compréhension de FL dans ce contexte : l'architecture des réseaux IoT qui en motive l'usage, les principes de ML et de DL sur lesquels il s'appuie, le protocole fédéré FL et ses principales variantes algorithmiques, ainsi que les techniques d'intelligence en essaim SI et d'optimisation évolutionnaire de plus en plus mobilisées pour traiter ses problèmes de conception combinatoire. La seconde adresse un état de l'art de FL robuste autour de quatre défis centraux : l'hétérogénéité statistique, la sélection des clients, le regroupement de modèles et les menaces de sécurité. Une discussion comparative réunit ces quatre axes sous un ensemble commun de critères d'évaluation et met en lumière des failles structurelles récurrentes dans la littérature, notamment un compromis persistant entre la robustesse aux données hétérogènes et la résilience face aux participants malveillants, une sous-utilisation des architectures hiérarchiques malgré leurs avantages, et l'absence des approches qui traitent la sélection des clients, le regroupement et l'agrégation comme un pipeline conçu de manière unifiée plutôt que comme des composantes indépendantes.