**People's Democratic Republic of Algeria Ministry of Higher Education and Scientific Research**

**University of Ferhat Abbas Setif I**

**Department of Computer Sciences**




A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

**Master's Degree in Cybersecurity , Ingénierie des Données et Technologie Web**

# Automated ADHD Detection Using Machine Learning, Deep Learning, and Transformer-Based Attention

Prepared and Presented by:

**IZOUNTAR Mohammed Said ZIET Salah-eddine**

**Supervisor:**

**Dr SEMCHEDINE Moussa Co-Supervisor:**

**Mem Bouhafs Amira**

**Promotion: 2025/2026**

Submitted on

**June, 2026**

# Abstract

Automated diagnosis and subtyping of Attention Deficit Hyperactivity Dis-order (ADHD) face critical computational challenges: multi-site heterogeneity, severe class imbalances, and temporal data leakage. This thesis addresses these limitations by developing and evaluating deep learning, multimodal fusion, and decentralized frameworks using two distinct modalities: 3D structural MRI (3D sMRI) and Electroencephalography (EEG).

For 3D sMRI (ADHD-200), a Multimodal Late-Fusion Attention Autoen-coder integrates volumetric scans with 74 tabular phenotypic features to achieve a 62.60% multi-class validation accuracy. Additionally, a hierarchical diag-nostic cascade decouples binary presence screening (62.00% accuracy) from granular subtyping (46.00% accuracy), effectively isolating the complex ADHD-Combined phenotype (0.33 F1-score). To preserve institutional data privacy, a decentralized Federated Learning framework (FedAvg) deployed across 9 clini-cal sites handles non-IID data distribution to yield a robust 58.59% global test accuracy.

For EEG functional streams (Shahed University), a patient-strict validation strategy partitioned by unique Subject IDs eliminates row-wise data leakage, ensuring a valid 80.00% patient-level accuracy. Concurrently, a hybrid 1D-CNN-LSTM network effectively models cross-channel topologies and long-range macro-temporal dependencies, reaching a 79.70% frame-level accuracy across 433,277 streaming segments.

Ultimately, synthesizing spatial-temporal deep architectures, multimodal phe-notypic late-fusion, and privacy-preserving decentralized networks provides a statistically rigorous, highly generalizable foundation for computer-aided neu-rodevelopmental diagnostics.

**Keywords:** Attention Deficit Hyperactivity Disorder (ADHD), Deep Learning, Federated Learning, Multimodal Data Fusion, Electroencephalography (EEG), Structural MRI (sMRI), Privacy Preservation.

# الملخص

يواجه التشخيص الا لي وتحديد الالأنواع الفرعية لالاضطراب فرط الحركة ونقص الالانتباه ADHD تحديات برمجية وحسابية كبيرة، مثل اختتلاف طبيعة البيانات بين المراك؛ الطبية، وعدم توازن أعداد الحالالات بين الف%ات، ومشكلة تسرب البيانات الزمنيc تعال; هذه الالأطروحة هذه المشكلالات بتطوير وتقييم نماذج تعتمد على التعلم العميق، ودجج البيانات المتعددة، والالأنظمة اللالامرك؛ ية، وذلك باستخدام نوعين من الفحوصات: التصوير بالرنين المغناطيسي البنيوي ثتلائي الالأبعاد SD sMRI وتخطيط كهربائية الدماغ EEGc

بالنسبة لبيانات الرنين المغناطيسي ،ADHD-200 تم تطوير نموذج مشفر تلقااي يعتمد على تقنية الالانتباه المتأخر والدجج Autoencoder Attention Late-Fusion Multimodal ا الصور لدجج'~ميزة 74 مع مية جدولية، ومكننا ذلك من تحقيق دقة تحقق بلغت c62%60 لتصنيف الحالالات المتعددةc بالإضافة إلى ذلك، تم تصميم نظام أشخيصي هريي يفصل بين التشخيص الثنااي )مصاب أو سليم( بدقة 62%، وتحديد الالأنواع الفرعية الدقيقة بدقة c46%00، ثما ساعد في تمييز الاصائص المعقدة للالاضطراب بمقياس F1-score بلغ cSSc0 و'ح7فاظ على خصوصية بيانات المرضى، تم تطبيق نظام التعلم اللالامرك؛ي FedAvg المشترك بين 9 مراك؛ طبية للتعامل مع التوز يع غير المتماثل للبيانات، وحقق ال:فوذج دقة اختبار عامة وقو ية بلغت c59%c58

أما بالنسبة لبيانات تخطيط كهربائية الدماغ EEG )من جامعة شاهد(، فقد تم تطبيق طر يقة تحقق صارمة تعتمد على معرفات المرضى لمنع تسرب البيانات بين عينات الالاختبار والتدريب، ثما صمن دقة حقيقية على مستوى المر يض بلغت c00%c80 وفي الوقت نفسه، تم استخدام شبكة مجينة تمع بين نماذج 1D-CNN وLSTM لتحليل الروابط بين قنوات الدماغ والتغيرات الزمنية طو يلة المدى بكفاءة، ثما ممح بالوصول إلى

دقة c79%70 على مستوى الإطارات الزمنية عبر أكثر من 4SS,000 شريحة بيانات متدفقةc

**الكلمات المفتاحية:** اضطراب فرط الحركة ونقص الالانتباه ،(ADHD) التعلم العميق، التعلم اللالامرك؛ي المشترك، دجج البيانات المتعددة، تخطيط كهربائية الدماغ ،(EEG) التصوير بالرنين المغناطيسي البنيوي ،(sMRI) حماية الاصوصيةc

# Résumé

Le diagnostic automatisé et le sous-typage du trouble déficitaire de l'attention avec hyperactivité (TDAH) se heurtent à des difficultés informatiques majeures : hétérogénéité multisite, déséquilibres importants entre les classes et fuites de données temporelles. Cette thèse s'attaque à ces limites en élaborant et en évaluant des cadres d'apprentissage profond, de fusion multimodale et décentralisés, à l'aide de deux modalités distinctes : l'IRM structurelle 3D (IRM-3D) et l'électroencéphalographie .(EEG)

Pour l'IRM-3D (TDAH-200), un auto-encodeur d'attention à fusion tardive multimodale intègre des acquisitions volumétriques à 74 caractéristiques phénotypiques tabulaires, atteignant une précision de validation multiclasse de 62,60 %. De plus, une cascade diagnostique hiérarchique dissocie le dépistage binaire de présence (précision de 62,00 %) du sous-typage granulaire (précision de 46,00 %), isolant efficacement le phénotype complexe du TDAH combiné (score F1 de 0,33). Afin de préserver la confidentialité des données institutionnelles, un cadre d'apprentissage fédéré décentralisé (FedAvg), déployé sur 9 sites cliniques, gère la distribution des données non-IID et .atteint une précision globale robuste de 58,59 % lors des tests

Pour les flux fonctionnels EEG (Université Shahed), une stratégie de validation stricte par patient, partitionnée par identifiants de sujet uniques, élimine les fuites de données ligne par ligne, garantissant une précision de 80,00 % au niveau du patient. Parallèlement, un réseau hybride 1D-CNN-LSTM modélise efficacement les topologies intercanaux et les dépendances macro-temporelles à long terme, atteignant une précision de 79,70 % au niveau de la trame sur 433 277 .segments de flux

En définitive, la synthèse d'architectures profondes spatio-temporelles, d'une fusion tardive phénotypique multimodale et de réseaux décentralisés préservant la confidentialité fournit une base statistiquement rigoureuse et hautement généralisable pour le diagnostic neurodéveloppemental .assisté par ordinateur



Mots-clés : Trouble du déficit de l’attention avec hyperactivité (TDAH), apprentissage profond, apprentissage fédéré, fusion de données multimodales, électroencéphalographie (EEG), IRM .structurelle (IRM-s), préservation de la vie privée

# Contents

Contents ..... iii
List of Figures ..... iv
List of Tables ..... v

**Introduction** **1**

**1 Theoretical foundation of ADHD and artificial intelligence** **3**
1.1 Understanding ADHD ..... 3
1.1.1 Definition and history ..... 3
1.1.2 Core clinical symptoms and presentations ..... 4
1.1.3 Etiology and causes ..... 4
1.2 The Neurobiology of ADHD ..... 5
1.2.1 Affected brain regions and structural morphometry ..... 5
1.2.2 Large-Scale crain networks ..... 5
1.2.3 Brain connectivity and the dysconnectivity hypothesis ..... 6
1.3 Traditional vs. Automated ADHD detection ..... 6
1.3.1 Current diagnostic tools ..... 6
1.3.2 Defining Automated ADHD Detection ..... 7
1.3.3 Advantages of AI-Based systems ..... 7
1.4 Data foundations ..... 9
1.4.1 Neuroimaging Modalities ..... 9
1.4.2 Comparison of Neuroimaging Modalities for AI ..... 9
1.5 Preprocessing ..... 9
1.5.1 Concept of preprocessing ..... 9
1.5.2 Key Preprocessing Steps ..... 10
1.5.3 Feature Extraction and Selection ..... 11
Conclusion ..... 12

**2 Theoretical Concept and Related Work** **13**
2.1 Machine Learning ..... 13
2.2 Types of Machine Learning ..... 13
2.2.1 Supervised Learning ..... 14
2.2.2 Unsupervised Learning ..... 14
2.2.3 Semi-Supervised Learning ..... 14
2.2.4 Reinforcement Learning ..... 15
2.3 Classical Machine Learning Algorithms ..... 15
2.3.1 Support Vector Machines (SVM) ..... 15
2.3.2 Decision Trees ..... 15

2.3.3 Random Forest.....15
2.3.4 XGBoost.....16
2.3.5 Comparative Analysis of Classical Algorithms.....17

2.3.6 Related Work . . . . . 17
2.4 Deep Learning . . . . . 18
2.4.1 Artificial Neural Networks (ANN) . . . . . 18
2.4.2 Convolutional Neural Networks (CNN) . . . . . 18
2.4.3 Long Short-Term Memory (LSTM) . . . . . 19
2.4.4 Related Work Using DL Models . . . . . 20
2.5 Autoencoder Architectures . . . . . 21
2.5.1 Core Components and Mathematical Framework . . . . . 21
2.5.2 Attention-based Autoencoders . . . . . 22
2.5.3 Application of Autoencoders in ADHD Detection . . . . . 22
2.6 Federated Learning . . . . . 22
2.6.1 FL Communication Protocols and Federated Averaging . . . . . 23
2.6.2 Privacy, Security, and Neuroimaging Applications . . . . . 23

**3 Methodology and Implementation** **25**
3.1 Development Environment and Software Frameworks . . . . . 25
3.1.1 Programming Languages and System Utilities . . . . . 25
3.1.2 Data Manipulation and Neuroimaging Frameworks . . . . . 25
3.1.3 Deep Learning and Machine Learning Engines . . . . . 26
3.1.4 Optimization Callbacks . . . . . 26
3.1.5 Visualization and Computational Platforms . . . . . 26
3.2 Dataset Description . . . . . 27
3.2.1 The ADHD-200 Consortium . . . . . 27
3.2.2 3D Structural MRI Technical Specifications . . . . . 27
3.2.3 Functional Electrophysiological Data Foundation . . . . . 27
3.3 Data Preprocessing and Spatial Cleaning . . . . . 28
3.3.1 Algorithmic Processing Pipeline . . . . . 28
3.3.2 3D Spatial Data Augmentation . . . . . 28
3.4 Proposed Models and Architectures . . . . . 29
3.4.1 Machine Learning Baseline Models . . . . . 29

3.4.2 Deep Learning Architectures . . . . . 31
3.4.3 Electroencephalogram (EEG) Functional Analysis Paradigms . . . . . 41
3.4.4 Regularization and Optimization Strategies (Overfitting Mitigation) . . 44
3.4.5 Network Regularization Techniques . . . . . 44
3.4.6 Optimization and Convergence Controls . . . . . 44
3.4.7 Model Evaluation Metrics . . . . . 45
3.4.8 The Confusion Matrix . . . . . 45
3.4.9 Performance Metrics . . . . . 45
3.5 Conclusion . . . . . 46

**4 Experiments, Results, and Discussion** **47**
4.1 Machine Learning Results . . . . . 47
4.1.1 Decision Tree Classifier . . . . . 47
4.1.2 Support Vector Machine (SVM) . . . . . 47
4.1.3 Random Forest Classifier. . . . . 48
4.1.4 XGBoost Classifier . . . . . 49
4.1.5 Overall Machine Learning Performance Summary . . . . . 50
4.2 Deep Learning Results . . . . . 52
4.2.1 LSTM Networks . . . . . 52
4.2.2 3D Convolutional Neural Networks . . . . . 52

4.2.3 Standard 3D Deep Neural Network Performance....53
4.2.4 Hybrid Spatial-Temporal Networks....54
4.2.5 Federated Learning Framework Performance....55
4.2.6 Attention-Based Autoencoders: Pure Imaging Performance....57
4.2.7 Hierarchical Diagnostic Cascades Performance....57
4.2.8 Attention-Based Autoencoders: Multimodal Late-Fusion Performance58
4.2.9 Electroencephalogram (EEG) Functional Analysis Performance....58
4.3 Comparative Performance Analysis....60
4.3.1 Performance Comparison Across All Proposed Models....60
4.3.2 Strengths and Weaknesses of Each Architecture....61
4.3.3 Ablation Studies and Key Architectural Insights....62
4.4 Comparison with Related Work....63
4.4.1 Comparison with Classical ML Studies....63
4.4.2 Comparison with Deep Learning and Transformer-Based Approaches64
4.4.3 Overall Positioning of the Current Work....65

**5 Conclusion and Future Work 67**
5.1 Concluding Remarks....67
5.2 Limitations of the Study....68
5.3 Future Work....68

# List of Figures


1.1 Figure 1.2 :Key neuroanatomical regions and functional networks targeted by Deep Learning..... 6
1.2 Figure 1.3: The Automated Detection Pipeline.....8

2.1 Figure 2.2: Taxonomy of primary machine learning paradigms [24].....14
2.2 Hierarchical structure of a standard decision tree model [51].....16
2.3 General CNN architecture for feature extraction and classification [71].....19
2.4 Internal architecture and gating mechanisms of an LSTM cell [78].....20
2.5 Standard autoencoder architecture and latent space reconstruction [87].....22

3.1 Architecture of the hybrid spatial-temporal 3D CNN-LSTM network.....33
3.2 Three-phase Federated Learning architecture from regional training to global evaluation. ..... 35
3.3 Structural pipeline of the 3D Attention Autoencoder framework.....37
3.4 Operational diagram of the hierarchical diagnostic cascade from binary to multi-class classification.....38
3.5 Architecture of the Multimodal Late-Fusion Attention Autoencoder framework.40
3.6 Phase 1 neural engineering pipeline using a 1D Convolutional Autoencoder. 42
3.7 Phase 2 validation framework with patient-strict subject segregation.....43
3.8 Phase 3 hybrid 1D-CNN-LSTM architecture for frame classification.....43

4.1 Feature Matrix Property and Distribution Analysis for Decision Tree Input Data. 48
4.2 Confusion Matrix for Decision Tree Classifier Across Clinical Phenotypes....48
4.3 Confusion Matrix for Support Vector Machine (SVM) Classifier Across Clinical Phenotypes. ..... 49
4.4 Feature Distribution Metrics for Random Forest Training Matrix.....49
4.5 Confusion Matrix for Random Forest Classifier Across Clinical Phenotypes. 50
4.6 Phenotypic Vector Matrix Properties for XGBoost Evaluation.....51
4.7 Confusion Matrix for XGBoost Classifier Across Clinical Phenotypes.....51
4.8 Multi-Class Confusion Matrix for Sequential LSTM Classification.....53
4.9 Training and Validation Accuracy/Loss Trajectories for the Stacked LSTM Net-work. 53
4.10 Training loss and accuracy curves for the 3D DNN model.....54

4.11 Confusion matrix for the standard 3D DNN model.....55
4.12 Multi-Class Confusion Matrix for the Hybrid 3D CNN-LSTM Architecture.56
4.13 Training and Validation Accuracy/Loss Trajectories for the Hybrid Model Pipeline. 56

# List of Tables

1.2.1 Historical Evolution of ADHD in Clinical Literature....4
1.4.1 Comparison Between Traditional Clinical Diagnosis and AI-Based Diagnosis for ADHD 7
1.4.3 Comparison of Modalities for AI-Based ADHD Detection....10

2.3.1 Comparison of SVM, Random Forest, and XGBoost in Neuroimaging Tasks [62, 64]....17

3.2.1 Demographic and phenotypic breakdown of the ADHD-200 sample....27
3.4.2 LSTM Network Architectural Layer Configuration....31
3.4.3 3D CNN Structural Details and Training Configuration....32
3.4.4 Standard 3D DNN Layer Specifications and Regularization Parameters....32
3.4.5 Hybrid 3D CNN-LSTM Network Layer Details and Sequence Formats....34
3.4.6 Decentralized Federated Learning Setup and Global Aggregation Parameters 36
3.4.7 Layer Specifications for the 3D Attention Autoencoder....38
3.4.8 Hierarchical Cascade Model Operational Configurations and Stage Splits....39
3.4.9 Multimodal Late-Fusion Attention Autoencoder Operational Hyperparameters . 40

4.1.1 Decision Tree Classifier Key Parameters and Accuracy....47
4.1.2 SVM Classifier Key Parameters and Accuracy....47
4.1.3 Random Forest Classifier Key Parameters and Accuracy....49
4.1.4 XGBoost Classifier Key Parameters and Accuracy....50
4.1.5 Performance metrics of baseline machine learning classifiers....51
4.2.6 Detailed LSTM Classifier Evaluation Matrix and Categorical Metrics....52
4.2.7 Volumetric 3D CNN Balanced Accuracy Convergence Log Over 15 Training Epochs 53
4.2.8 Classification report and global performance metrics for the standard 3D DNN. 54
4.2.9 Classification performance of the hybrid 3D CNN-LSTM model....55
4.2.10.... Detailed Diagnostic Classification Report for the Unified FedAvg Network....56
4.2.11 3D Attention Autoencoder Multi-Task Training and Validation Optimization Log 57
4.2.12 Empirical Metric Log for the Two-Stage Hierarchical Diagnostic Cascade .... 58
4.2.13 Patient-Isolated Classical Ensemble Diagnostic Performance Distribution....59
4.2.14 Deep Hybrid 1D-CNN-LSTM Frame-Level Classification Metrics....59

4.3.15Consolidated Master Performance Comparison Across Architectural Frameworks 60
4.4.16Systemic Comparison and Positioning Matrix Against Contemporary Literature 65