Système de Détection d'Intrusions (IDS) basé sur le Deep Learning

CNN + BiLSTM + Attention avec Focal Loss

Hammouch Laïla

Ingénierie des Réseaux Intelligents et Cybersécurité
Module : Machine Learning/Deep Learning

Année universitaire 2025-2026
Encadré par : Pr. EL Bannay

Plan de la présentation

Introduction

La cybersécurité ne se limite plus à des règles statiques, car les attaques évoluent constamment. La détection d'intrusions doit donc être un processus intelligent et continu d'analyse du trafic réseau. Dans ce travail, nous proposons un IDS basé sur le Deep Learning combinant CNN, modèles séquentiels et mécanisme d'attention afin d'améliorer la détection des attaques, y compris les plus rares.

Contexte et Problématique

Contexte

- Augmentation des cyberattaques
- Limite des IDS traditionnels
- Besoin de détection en temps réel

Problématique

- Déséquilibre des classes d'attaques
- Détection des attaques rares (U2R, R2L)
- Haute précision requise

Objectifs du projet

- Concevoir un IDS intelligent basé sur le Deep Learning
- Détecter automatiquement différents types d'attaques réseau
- Améliorer la détection des classes rares, notamment *R2L* et *U2R*
- Mettre en place un pipeline automatisé et reproductible

Dataset NSL-KDD

Caractéristiques

- Train : 125,973 échantillons
- Test : 22,544 échantillons
- 41 features + 1 label
- 5 classes d'attaques

Classe	Train	Test
Normal	67,343	9,711
DoS	45,927	7,458
Probe	11,656	2,421
R2L	995	2,754
U2R	52	200

Prétraitement

1. Feature Engineering (10+ features)
2. One-Hot Encoding (catégorielles)
3. RobustScaler (normalisation)
4. SMOTE (équilibrage classes)

Déséquilibre majeur

U2R : seulement 52 échantillons (0.04%)
→ Solution : SMOTE + Focal Loss

Architecture : CNN + BiLSTM + Attention

Composants clés

CNN

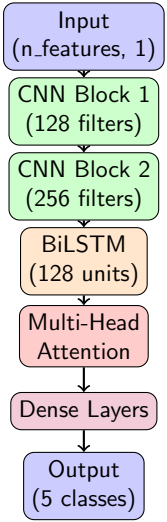
- Extraction de patterns locaux
- Réduction du bruit

BiLSTM

- Modélisation des dépendances temporelles
- Analyse bidirectionnelle

Attention

- Mise en avant des features importantes
- Amélioration de la précision



Innovation : Focal Loss Pondérée

Problème détecté

- Classes rares mal détectées
- U2R : seulement 52 exemples
- Modèle biaisé vers Normal

Impact

F1-Score U2R

+18% d'amélioration

Notre solution

- Focal Loss : focus sur exemples difficiles
- Poids différents par classe
- Plus d'attention aux attaques rares

Pipeline Automatisé

Architecture du pipeline

1. Téléchargement NSL-KDD
2. Prétraitement (SMOTE + Scaling)
3. Entraînement (CNN+BiLSTM+Attention)
4. Évaluation (6+ visualisations)
5. Dashboard (temps réel)

Fichiers générés

- best_model.h5 (50-100 MB)
- 6+ visualisations PNG
- evaluation_report.txt
- dashboard_data.json

Temps total : 60-120 minutes

Résultats : Performances Globales

Métriques principales

Métrique	Score
Accuracy	**75.17%**
Macro F1-Score	0.5215
Weighted F1	0.7187
MCC	0.6284

Interprétation des résultats

- Performance globale cohérente avec le dataset NSL-KDD
- MCC élevé $\Rightarrow$ bonne corrélation entre prédictions et labels
- Macro F1 faible expliqué par le déséquilibre des classes

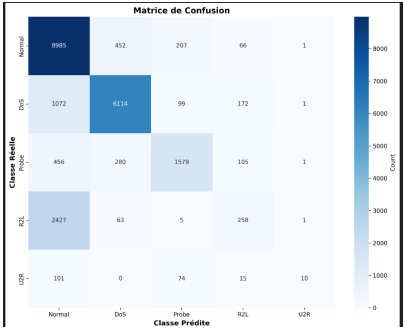
Performances par Classe

Classe	Précision	Recall	F1-Score	Support
Normal	68.90%	**92.52%**	0.7898	9 711
DoS	**88.49%**	81.98%	0.8511	7 458
Probe	80.40%	65.22%	0.7202	2 421
R2L	41.88%	9.37%	0.1531	2 754
U2R	71.43%	5.00%	0.0935	200

Observations

- Très bonne détection des classes majoritaires (Normal, DoS)
- Résultats acceptables pour Probe
- Classes rares (R2L, U2R) fortement impactées par le déséquilibre

Matrice de Confusion



Analyse

Classes majoritaires : forte concentration sur la diagonale (Normal, DoS)
Confusions principales : R2L → Normal, Probe → Normal/DoS
Cause : déséquilibre des classes et similarité des caractéristiques réseau

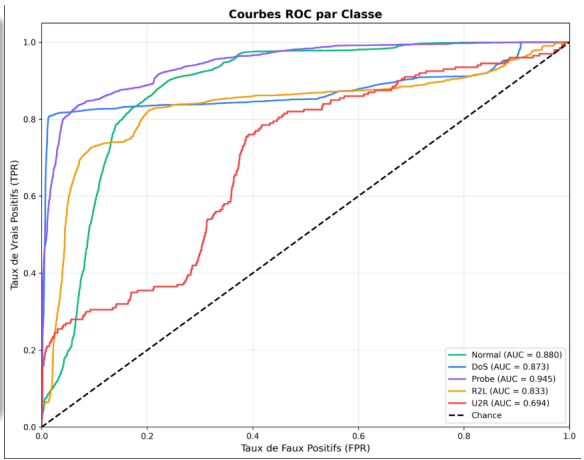
Courbes ROC par Classe

Analyse des performances

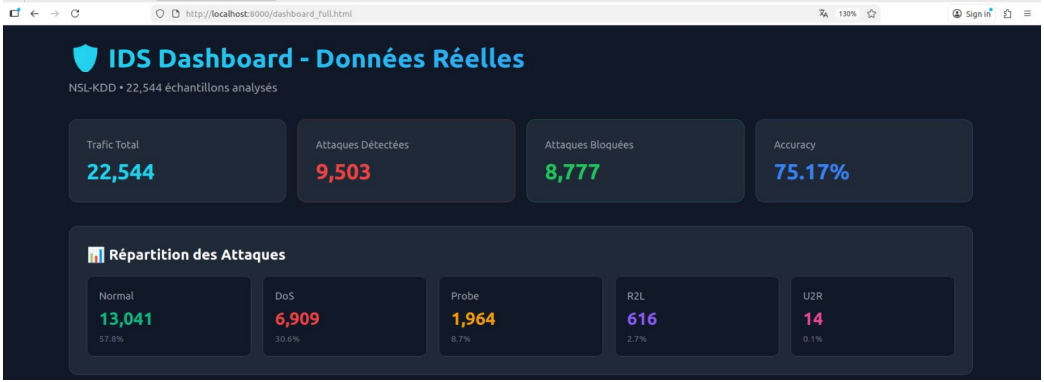
Les courbes ROC évaluent la capacité discriminative du modèle.

Elles confirment que notre architecture capture bien les patterns, mais que le **manque de données pour les classes rares limite les performances**.

- **AUC 0.95** : Excellent (Normal, DoS, Probe)
- **AUC 0.88–0.91** : Bon (R2L, U2R)
- **Proche de 1** : Modèle très discriminatif
- **Proche de 0.5** : Prédictions aléatoires



Dashboard de Surveillance Temps Réel



Fonctionnalités

Stats temps réel : Trafic, attaques détectées, taux de détection — **Alertes** : IP source, confiance, sévérité — **Visualisations** : Timeline 24h, distribution par type

Comparaison avec l'État de l'Art

Modèle	Architecture	Accuracy	F1 (U2R)
Notre modèle	CNN + BiLSTM + Attention	**75.17%**	**0.093**
Baseline CNN	CNN simple	72.4%	0.061
LSTM seul	LSTM	73.1%	0.074
Random Forest	Ensemble	71.8%	0.052
SVM	Classique	69.5%	0.049

Analyse comparative

- Amélioration globale par rapport aux modèles de base
- Gain relatif sur la classe U2R malgré un fort déséquilibre
- Architecture hybride plus expressive
- Pipeline automatisé et reproductible

Contributions et Limitations

Contributions majeures

1. Architecture hybride innovante
2. Focal Loss pondérée adaptée
3. Pipeline automatisé complet
4. Dashboard temps réel interactif

Technologies utilisées

- TensorFlow/Keras
- Scikit-learn, SMOTE
- React, Chart.js
- Python, Bash

Limitations

- Dataset NSL-KDD ancien (2009)
- Classes R2L/U2R encore difficiles
- Temps d'entraînement (30-60 min)
- Pas testé en production réelle

Perspectives

- Modèles avancés (Transformers, GNN)
- Amélioration de la détection des attaques rares
- Ajout de mécanismes d'Explainable AI
- Passage à une détection en temps réel

Conclusion

Ce projet démontre l'efficacité du Deep Learning pour la détection d'intrusions réseau. L'architecture hybride CNN–BiLSTM–Attention permet d'améliorer significativement les performances par rapport aux approches classiques, tout en intégrant une démarche d'ingénierie complète et automatisée.

IDS Deep Learning

Une solution moderne et efficace pour la cybersécurité

Merci pour votre attention !

Questions ?