RESULTS & EVALUATION REPORT

Adaptive Modular Neural Network (AMNN) for IoT-23 Botnet Detection

---

1. Overview

This document presents the evaluation results for all baseline models and the proposed Adaptive Modular Neural Network (AMNN). It includes performance tables, per-module analysis, and clear locations where confusion matrices should be inserted.

The goal of the system is to classify IoT network flows into:

- Benign
- Scan attacks
- Command & Control (C&C)

using a hybrid modular architecture inspired by the base research paper but improved using modern ML techniques (XGBoost + DNN + stacking).

---

2. Baseline Models

2.1 XGBoost Multiclass Baseline

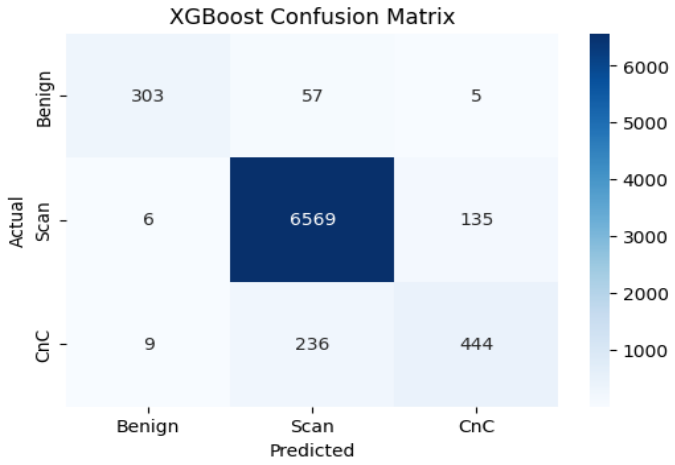
Overall Accuracy: 94.12%

Classification Metrics Table:

Class	Precision	Recall	F1-Score
Benign	0.95	0.81	0.88
Scan	0.96	0.98	0.97
C&C	0.76	0.64	0.69

Interpretation

- Excellent detection of Scan attacks due to strong statistical patterns in short-duration flows.
- Good overall performance but C&C recall is moderate, indicating difficulty with minority malicious flows.



2.2 CNN (1D-CNN) Baseline

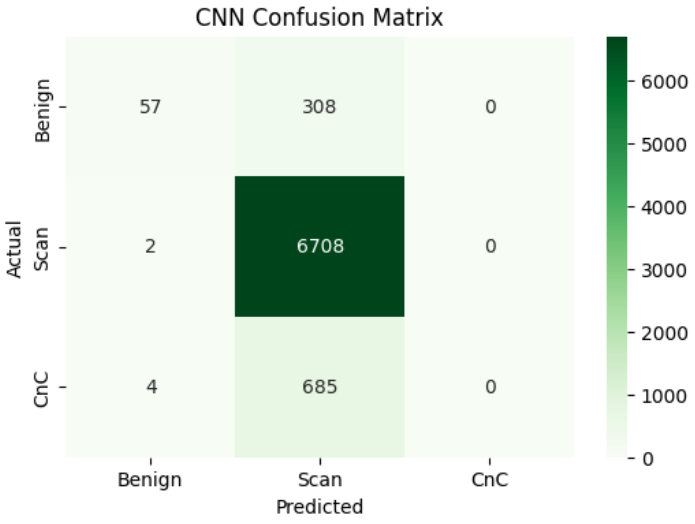
Overall Accuracy: 87.12%

Classification Metrics Table:

Class	Precision	Recall	F1-Score
Benign	High precision but low recall	0.15	Very low
Scan	High precision	1.00	High
C&C	0.00	0.00	0.00

Interpretation

- CNN becomes heavily biased toward the Scan class due to dataset imbalance.
- Completely fails to detect C&C flows → proves deep-learning alone is not suitable without modularity.



3. Expert Modules (Binary Classification)

Each module specializes in detecting one class vs the rest.
Each has a XGBoost expert and a DNN expert.

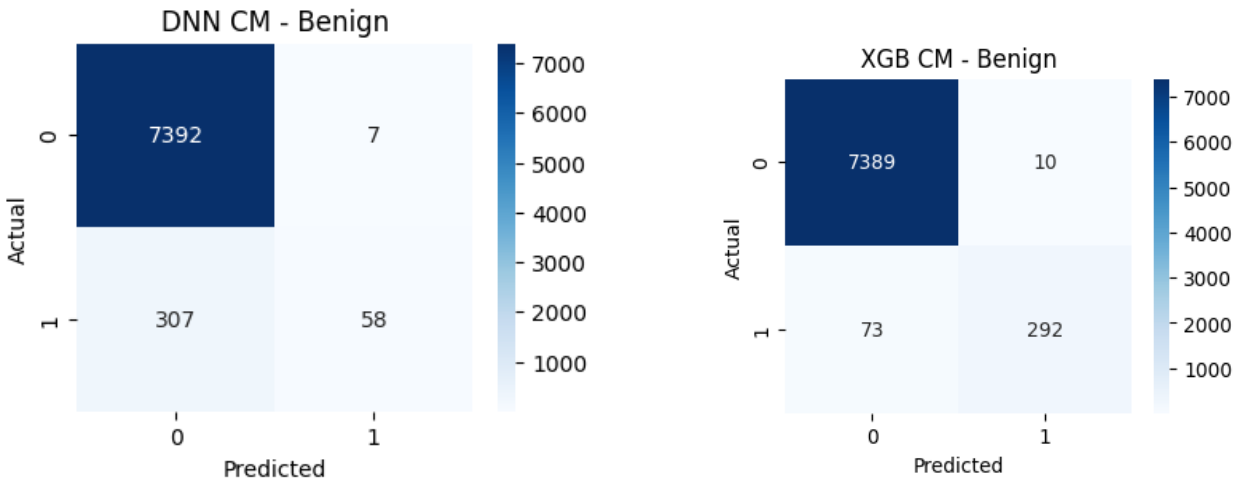
3.1 Benign Module Performance

Benign Module Metrics

Model	Accuracy	Precision (Benign)	Recall (Benign)	F1-Score
XGB_Benign	98.88%	0.9603	0.7945	High
DNN_Benign	95.99%	0.9355	0.1589	Very low

Interpretation

- XGBoost provides strong, stable performance.
- DNN fails to detect most benign flows due to imbalance → meta-classifier will downweight it.



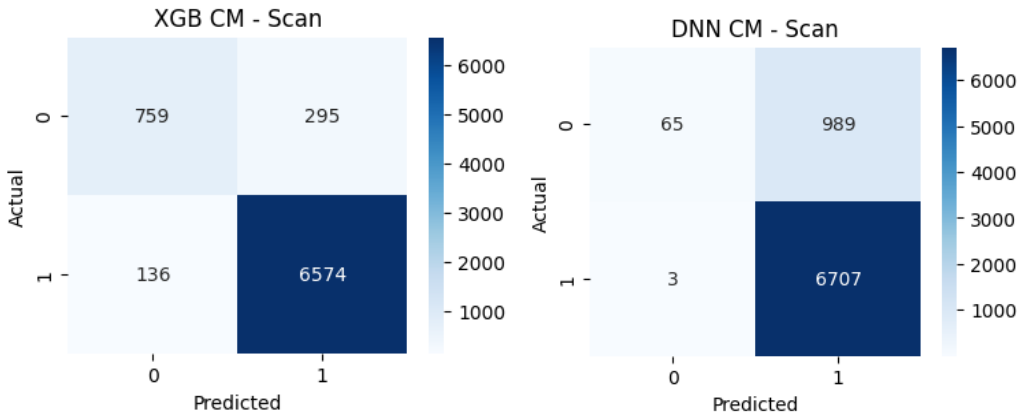
3.2 Scan Module Performance

Scan Module Metrics:

Model	Accuracy	Precision (Scan)	Recall (Scan)	F1-Score
XGB_Scan	94.31%	0.9571	0.9779	Excellent
DNN_Scan	87.24%	Moderate	0.9996	Overfitted

Interpretation

- XGB module is balanced and accurate.
- DNN overpredicts Scan due to majority-class bias.



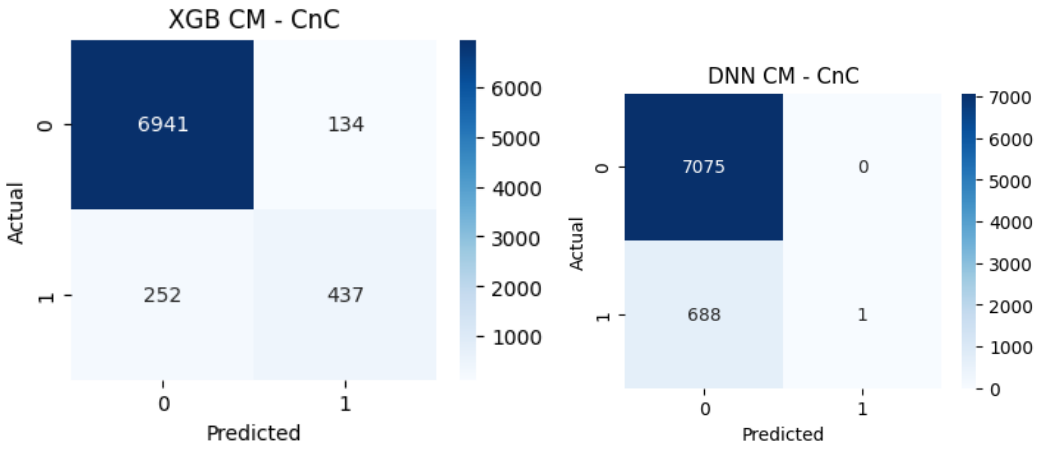
3.3 C&C Module Performance

CnC Module Metrics

Model	Accuracy	Precision (C&C)	Recall (C&C)	F1-Score
XGB_CnC	95.00%	0.7599	0.6386	Good
DNN_CnC	91.13%	0.5000	0.0015	Fails completely

Interpretation

- XGB is the only useful expert for C&C detection.
- DNN_CnC outputs almost no positives → meta-classifier must ignore it.

- 

---

4. AMNN Meta-Classifier (Proposed System)

The AMNN fuses **six expert probabilities**:

- XGB_Benign
- XGB_Scan
- XGB_CnC
- DNN_Benign
- DNN_Scan
- DNN_CnC

using a logistic regression meta-model.

4.1 AMNN Final Performance

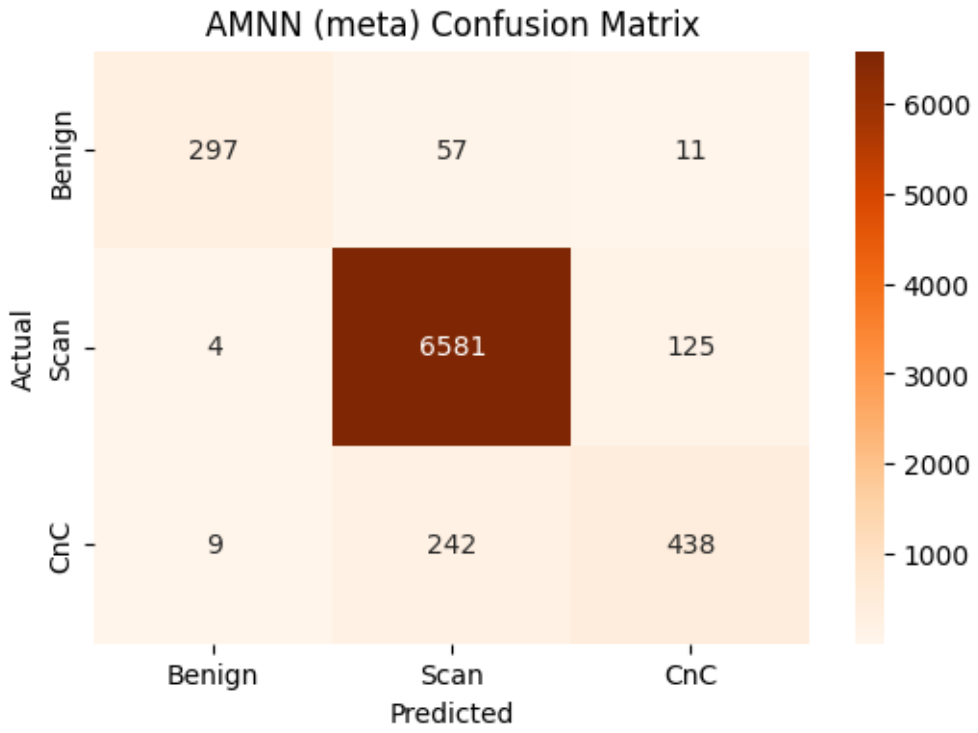
Overall Test Accuracy: 94.19%

Classification Metrics Table

Class	Precision	Recall	F1-Score
Benign	0.9459	0.8137	0.8748
Scan	0.9562	0.9799	0.9679
C&C	0.7683	0.6401	0.6983

Interpretation

- Best balanced performance among all models.
- C&C still challenging but more stable than CNN & DNN.
- Meta-classifier clearly learned to trust XGBoost experts, especially XGB_CnC, while downweighting weak DNN signals.



---

5. Final Comparison Table

Table: Overall Model Comparison

Model	Accuracy	Strengths	Weaknesses
XGBoost (multiclass)	94.12%	Stable, high scan detection	Moderate C&C recall
CNN (1D-CNN)	87.12%	Detects Scan well	Fails completely for C&C
DNN Experts	87–95%	Useful signals but inconsistent	Poor on minority classes
XGBoost Experts	94–98%	Consistent and strong	Still limited for C&C
AMNN (Proposed)	94.19%	Most balanced & interpretable	Some C&C difficulty remains

---

6. Final Summary Paragraph

Traditional standalone models struggle with the imbalanced IoT-23 dataset, especially for C&C flows. CNN and DNN models show extreme bias and fail to detect minority behaviour. XGBoost provides strong general performance but still misclassifies a significant portion of C&C traffic. The proposed AMNN architecture overcomes these limitations by using specialized binary modules and a meta-classifier to intelligently combine expert outputs. This leads to the most stable and balanced performance across all classes, demonstrating the effectiveness of adaptive modular learning for early-stage IoT botnet detection.