Multi-Modal Detection of Illicit Wildlife Trafficking Actors: A Hybrid AI Framework for Classification, Network Analysis, and Web Retrieval

Yan Pan Chung Stefano Scaini Shuo Yang

September 30, 2025

Abstract

This paper presents the methodology and results of our finalist submission to the 2023-2024 AI IMI Big Data Hub Competition. We propose a multi-stage machine learning pipeline to detect and understand illicit wildlife trafficking (IWT) actors using a combination of supervised classification, graph-based community detection, and retrieval-augmented generation (RAG). Our results demonstrate high accuracy in classification (F1 = 0.8), robust detection of suspicious pathways in transaction networks using GCNs (F1 = 0.85), and promising results in identifying traffickers from web data using a corrective RAG pipeline.

1 Introduction

Wildlife trafficking poses a critical threat to global biodiversity and security. Detecting the individuals and networks behind this illicit activity requires advanced analytical methods capable of integrating heterogeneous data sources. This project addresses the problem using three approaches: supervised learning for binary fraud classification, graph convolutional networks (GCN) for community-level pattern recognition, and large language models (LLMs) via RAG pipelines for open-source intelligence mining.

2 Task 1: Supervised Learning

2.1 Data Engineering and Preprocessing

Features were extracted from four main dataframes: KYC, Cash, EMT, and Wire. Categorical features like gender and occupation were one-hot encoded, while numeric features were normalized using StandardScaler. Significant class imbalance was observed (˜34x), with 3,543 positive samples among 127,262 total observations.

2.2 Resampling Strategies

We explored downsampling, upsampling, and a novel random up/down ensemble approach to handle imbalance. Agglomerative clustering was also attempted to partition the majority class into 36 clusters, but this underperformed (F1 = 0.03).

2.3 Model Performance

The best-performing method was an ensemble of MLP classifiers trained on randomly up/down sampled subsets. This achieved an F1 score of 0.80, outperforming logistic regression and random forest baselines.

2.4 Feature Importance

Feature permutation using a fully-connected neural net showed robust F1 deterioration for top features, enabling informed selection. Base model F1 = 0.76.

3 Task 2: GCN and Community Detection

3.1 Graph Construction

Nodes represented customer IDs with attributes from KYC, cash, EMT, and wire data. Edges were built using transaction meta-data, keyword filtering from message content, and transaction types.

3.2 GCN Model

We trained a two-layer GCN with tanh activations and BCE loss over 650 epochs. The model achieved an F1 score of 0.85 and predicted ˜5,400 suspicious nodes (1.8% of the graph).

3.3 Suspicious Pathway Analysis

We analyzed shortest paths between high-betweenness suspicious actors and found cliques with similar message scores, wire amounts, and transaction types.

3.4 Community Detection

Louvain modularity revealed 19 communities, 12 of which had high message scores. We filtered the graph to a 5-hop neighborhood of labeled nodes (reducing size by 44%) and applied spectral clustering on a dissimilarity matrix with connectivity constraints, yielding ˜550 clusters. Final clusters were analyzed using GNN scores to identify ˜50 likely trafficking hubs.

4 Task 3: Retrieval-Augmented Generation and LLM

4.1 Initial Attempts and Challenges

Initial attempts using LLM agents with toolchains proved unstable due to prompt sensitivity and inconsistent output.

4.2 Corrective RAG (CRAG) Pipeline

We implemented a Corrective RAG pipeline using Scrapy for scraping sources like justice.gov, Toronto Star, and CTV News. Chunks were embedded using MiniLM and stored in a Pinecone vector store. A LangChain-based agent used a grading prompt to evaluate top-5 retrieved documents for relevance. If any document was deemed irrelevant, the pipeline triggered a DuckDuckGo search to retrieve additional context before final decision-making.

4.3 Limitations and Outlook

While effective in principle, hallucination and poor understanding of subtle semantic distinctions (e.g., “animal trafficking” vs. “wildlife trafficking”) remain unresolved.

5 Results and Discussion

Each task produced actionable results:

- **Task 1:** F1 = 0.80 using MLP ensemble on up/down samples.
- **Task 2:** F1 = 0.85 with GCN, ˜50 trafficking clusters identified.
- **Task 3:** Proof-of-concept for CRAG with web retrieval for named traffickers.

Tradeoffs were evident between precision and recall, graph resolution and complexity, and LLM retrieval accuracy. Multi-modal fusion of all three approaches may yield even stronger performance in real-world enforcement settings.

6 Conclusion

We present a full-stack AI pipeline combining supervised learning, network analysis, and LLM-powered retrieval to identify actors involved in wildlife trafficking. Our hybrid approach proved effective across structured data, graph representations, and open-web resources. Future work includes hyperparameter optimization, improved interpretability for GCNs, and larger-scale web data ingestion.

References

- Haas, T.C. (2015). *Using Social Network Analysis to Disrupt Transnational Wildlife Trafficking Networks.*
- Keskin, B.K. (2021). *Quantitative Investigation of Wildlife Trafficking Supply Chains: A Review.*
- Shi-Qi et al. (2024). *Corrective Retrieval Augmented Generation.*
- Kipf, T.N. & Welling, M. (2017). *Semi-Supervised Classification with Graph Convolutional Networks.*