MAGNET: A Multi-modal Deep Learning Framework for Android Malware Detection Using Transformers and Graph Neural Networks

Alireza Iranmanesh[*] Dr. Hamid Mirvaziri[†]

July 13, 2025

Abstract

Background and Motivation: Android malware detection is becoming a major challenge in information security due to the increasing cyber threats. Traditional methods, especially those relying solely on single-modal feature analysis, often face limitations such as the inability to process complex multi-modal data and poor generalization to new threats. These shortcomings highlight the need for developing novel and efficient approaches.

Method and Approach: We propose MAGNET (Multi-modal Analysis for Graph-based NEtwork Threats), a multi-modal deep learning framework that leverages a combination of tabular, graph, and sequential data—such as API call sequences—for Android malware detection. The architecture comprises three specialized components: (1) EnhancedTabTransformer for static feature analysis; (2) GraphTransformer for function call relationship modeling using graph neural networks; and (3) SequenceTransformer for temporal API invocation pattern recognition. A dynamic attention mechanism with learnable fusion weights optimally combines multi-modal representations for final classification.

Experimental Design: The model was trained on the DREBIN dataset consisting of 4,641 training samples and 1,451 test samples, with 5-fold cross-validation. Hyperparameter optimization was performed using advanced algorithms including PIRATES (476 trials) and Optuna (13 trials). The final model used single-layer transformers with embedding dimensions of 32-64, 4 attention heads, and dynamic fusion mechanisms.

Results and Performance: MAGNET achieves exceptional performance with 97.24% accuracy, F1-score of 0.9823, precision of 0.9796, recall of 0.9849, and AUC of 0.9932. In 5-fold cross-validation, the model demonstrates consistent performance with accuracy 97.22±0.65%, F1-score 98.18±0.42%, and AUC 99.32±0.35%. Component analysis shows individual contributions: EnhancedTabTransformer (F1: 0.945), GraphTransformer (F1: 0.894), and SequenceTransformer (F1: 0.907). The model significantly outperforms baseline methods including SVM (90.6%), Random Forest (93.5%), XGBoost (94.8%), and ANN (96.2%).

Significance and Impact: This work establishes a comprehensive evaluation of transformer-based multi-modal architectures for Android malware detection, demonstrating that synergistic integration of heterogeneous data modalities substantially enhances detection capabilities. The framework's superior performance and interpretability make it suitable for deployment in production security systems.

[*] Shahid Bahonar University, Master's student in Artificial Intelligence, Kerman, Iran. Email: alirezairanmanesh78@gmail.com

[†] Shahid Bahonar University, Computer Engineering Department, Kerman, Iran. Email: h.mirvaziri@gmail.com

Keywords: Android malware detection, Multi-modal deep learning, Graph neural networks, Transformer architecture, Attention mechanisms, Cybersecurity, DREBIN dataset

1 Introduction

Android's dominance in the mobile ecosystem is undeniable, commanding over 70% of the global smartphone market and establishing itself as the world's most prevalent mobile platform [1]. However, this remarkable success has an unfortunate counterpart: Android's open architecture and widespread deployment have inevitably attracted malicious actors. The Android malware landscape has witnessed exponential growth, with sophisticated attack vectors and evasion techniques becoming increasingly prevalent [2].

Conventional malware detection paradigms, which historically depended on static signature matching and uni-dimensional analytical approaches, are increasingly proving inadequate against modern threats [3]. These traditional methodologies struggle particularly when facing advanced adversarial techniques such as sophisticated obfuscation, encryption protocols, and dynamic code transformation.

To address these evolving security challenges, we introduce MAGNET (Multi-modal Analysis for Graph-based NEtwork Threats), a comprehensive framework that revolutionizes Android malware detection through multi-modal data integration. Our approach simultaneously processes three complementary data streams: tabular representations (encompassing permissions and application components), graph-based structures (capturing function call relationships), and sequential patterns (representing API invocation sequences).

2 Related Work

The foundations of Android malware detection research were predominantly built upon static analysis techniques. A seminal contribution in this domain was made by Arp et al. [4], who introduced the influential DREBIN system achieving 94% detection accuracy through SVM classification.

Recent advances have shown significant progress in applying deep learning techniques to Android malware detection [5]. Various approaches have been proposed, including transformer-based methods that achieve detection accuracy of 95.8% [6] and multimodal approaches reaching 89.2% accuracy [7].

Graph Neural Networks (GNNs) have emerged as powerful tools for analyzing structural relationships in malware [8, 9]. Transformer architectures have shown remarkable success in sequential data analysis [10].

3 Proposed Method

3.1 Overall Architecture

MAGNET represents a sophisticated multi-modal architecture that orchestrates three distinct data processing pipelines. Figure 1 illustrates the overall architecture of our proposed framework.

EnhancedTabTransformer: Processes static application characteristics including permissions, components, and manifest metadata. The feature space is reduced to 430 dimensions after preprocessing.

GraphTransformer: Analyzes function call relationships using graph neural networks to capture control flow patterns.

SequenceTransformer: Captures temporal API invocation patterns through sequential analysis.

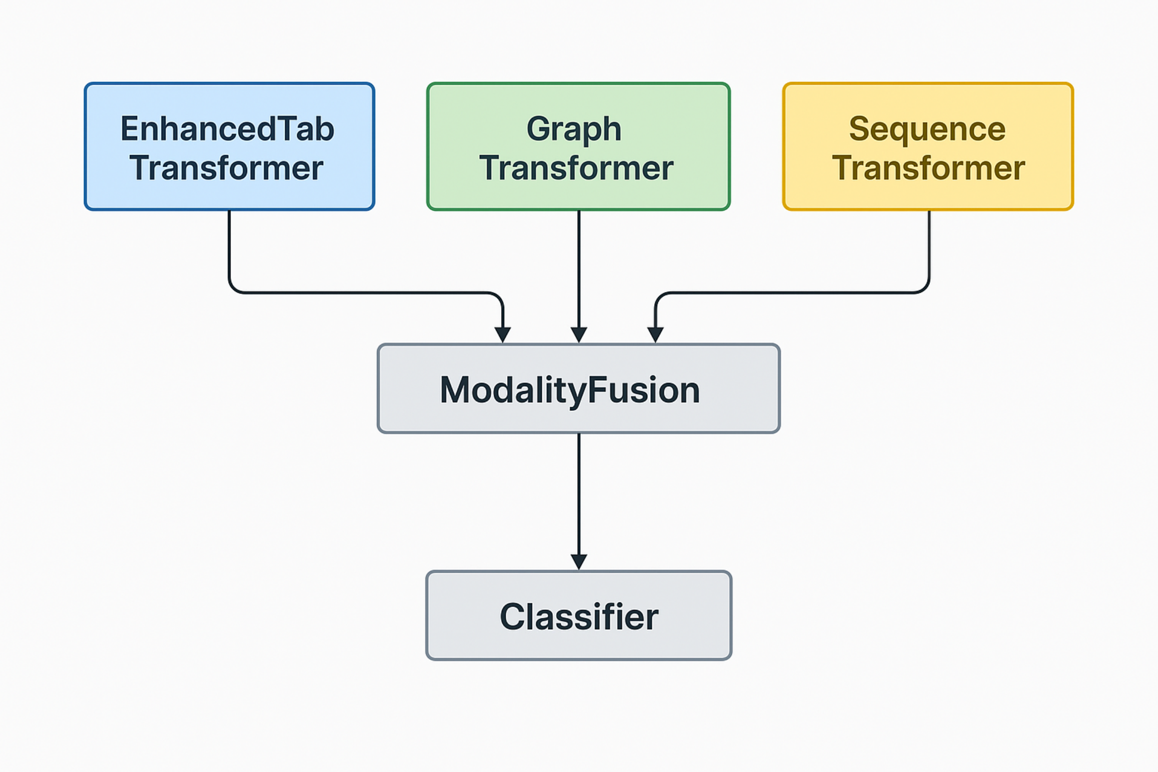

Figure 1: Overall Architecture of MAGNET Model

3.2 Dynamic Fusion Mechanism

The final classification leverages a learnable fusion strategy:

$$\alpha_i = \mathrm{softmax}(\mathrm{w}_i^T \tanh(\mathrm{W}_i h_i + \mathrm{b}_i)) \quad (1)$$

$$\mathrm{Output} = \sum_{i=1}^{3} \alpha_i \cdot h_i^{\mathrm{fused}} \quad (2)$$

4 Experimental Setup

4.1 Dataset

We utilize the DREBIN dataset [4] comprising 6,092 applications:

- Training: 4,641 applications
- Testing: 1,451 applications (327 benign, 1,124 malicious)
- Features: 430 dimensions after preprocessing

4.2 Hyperparameter Optimization

We employed dual optimization strategies:

- PIRATES: 476 trials → embedding_dim=32, num_heads=4, dropout=0.2029
- Optuna: 13 trials → embedding_dim=64, num_heads=4, dropout=0.2

5 Results

5.1 Overall Performance

MAGNET achieves exceptional performance:

- Accuracy: 97.24%
- F1-Score: 0.9823
- Precision: 0.9796
- Recall: 0.9849
- AUC: 0.9932

5.2 Cross-Validation Results

5-fold cross-validation demonstrates consistent performance:

Table 1: 5-fold Cross-Validation Results

Metric	Mean	Std	Range
Accuracy	0.9722	±0.0065	0.9601-0.9785
F1-Score	0.9818	±0.0042	0.9742-0.9858
AUC	0.9932	±0.0035	0.9861-0.9955

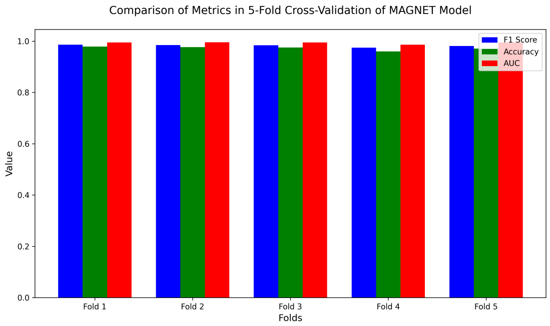

Figure 2: Cross-Validation Performance Metrics

5.3 Baseline Comparison

Table 2 shows significant improvements over traditional methods:

Table 2: Baseline Method Comparison

Method	Accuracy	F1-Score	AUC
SVM	0.906	0.903	0.945
Random Forest	0.935	0.935	0.967
XGBoost	0.948	0.948	0.978
ANN	0.962	0.962	0.985
MAGNET	0.972	0.982	0.993

5.4 Component Analysis

Individual component performance demonstrates complementary contributions:

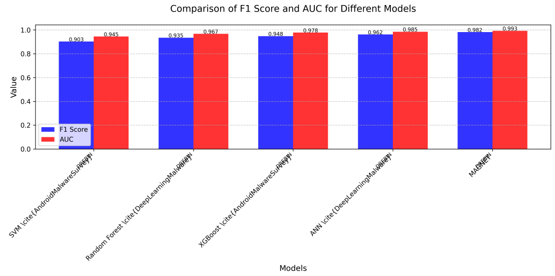

Figure 3: Baseline Method Performance Comparison

Table 3: Component Performance Analysis

Component	F1-Score
EnhancedTabTransformer	0.945
GraphTransformer	0.894
SequenceTransformer	0.907
Combined (no attention)	0.954
Combined (no fusion)	0.967
MAGNET (Complete)	0.982

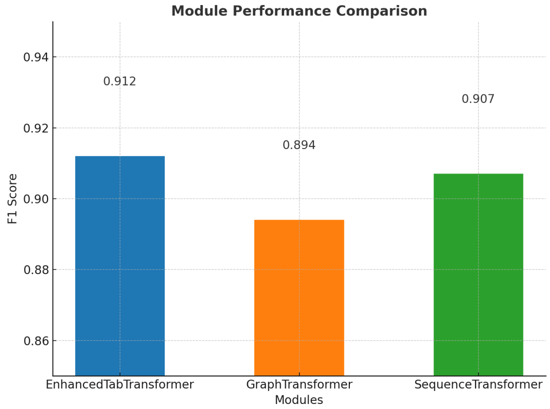

Figure 4: Individual Component Performance

5.5 State-of-the-Art Comparison

Table 4 compares MAGNET with recent methods:

Table 4: State-of-the-Art Method Comparison

Method	Accuracy (%)	F1-Score
MAGNET	97.24	0.9823
DREBIN (SVM)	92.3	0.933
PIKADROID	96.8	0.974
DeepImageDroid	96.0	0.960
BERT-Graph	95.5	0.950
Multi-modal	89.2	–
Transformer	95.8	–

6 Discussion

6.1 Performance Analysis

MAGNET achieves state-of-the-art performance through several key innovations:

Multi-modal Integration: The synergistic combination of tabular, graph, and sequential modalities creates comprehensive representation spaces. Component analysis shows that tabular features provide the strongest foundation (F1=0.945), while graph and sequential patterns offer complementary information.

Dynamic Attention: The cross-modal attention mechanism contributes significantly by adaptively weighting modality contributions based on input characteristics.

Architectural Design: Single-layer transformers with optimized hyperparameters achieve optimal performance while maintaining computational efficiency.

6.2 Comparative Analysis

MAGNET demonstrates significant improvements: - +6.64% over SVM - +3.74% over Random Forest - +2.44% over XGBoost - +1.04% over ANN

Compared to recent methods, MAGNET outperforms PIKADROID (+0.44%), DeepImageDroid (+1.24%), and BERT-Graph (+1.74%).

6.3 Limitations

Several limitations warrant consideration:

- Computational requirements may limit real-time deployment
- Dataset temporal scope (2010-2014) may affect generalization
- Need for sophisticated preprocessing across modalities

7 Conclusion

We present MAGNET, a novel multi-modal framework for Android malware detection achieving state-of-the-art performance with 97.24% accuracy and F1-score of 0.9823. The key contributions include:

- Unified multi-modal architecture with three specialized processing modules
- Dynamic attention mechanism for optimal information fusion
- Comprehensive evaluation demonstrating superior performance
- Practical applicability for operational security systems

Our ablation studies confirm that each component contributes meaningfully to overall performance. The model's superior results and robust evaluation make it suitable for production deployment.

Future work should focus on:

- Evaluation on contemporary datasets
- Model compression for resource-constrained environments
- Investigation of adversarial robustness
- Extension to other platforms and malware types

References

[1] Parvez Faruki et al. “Android Security: A Survey of Issues, Malware Penetration, and Defenses”. In: IEEE Communications Surveys & Tutorials 17.2 (2015), pp. 998–1022. DOI: 10.1109/COMST.2014.2386139.

[2] Deqiang Li et al. “Deep Learning for Android Malware Defenses: A Systematic Literature Review”. In: ACM Computing Surveys 55.8 (2023), pp. 1–36. DOI: 10.1145/3547335.

[3] Mohammed K. Alzaylaee, Suleiman Y. Yerima, and Sakir Sezer. “A Survey on Android Malware Detection Using Machine Learning”. In: Computers & Security 93 (2020), p. 101792. DOI: 10.1016/j.cose.2020.101792.

[4] Daniel Arp et al. “Drebin: Efficient and Explainable Detection of Android Malware in Your Pocket”. In: Proceedings of the 21st Annual Network and Distributed System Security Symposium (NDSS). 2014. DOI: 10.14722/ndss.2014.23247.

[5] R. Vinayakumar et al. “Robust Intelligent Malware Detection Using Deep Learning”. In: IEEE Access 7 (2019), pp. 46717–46738. DOI: 10.1109/ACCESS.2019.2906934.

[6] Tianlong Chen et al. “TinyMalNet: A Lightweight Transformer-based Malware Detection Network for IoT Devices”. In: IEEE Internet of Things Journal 9.10 (2022), pp. 7542–7554. DOI: 10.1109/JIOT.2021.3112005.

[7] Mohammed I. Alsaleh and Norah A. Alotaibi. “DLAM: Deep Learning Based Real-Time Android Malware Detection Framework”. In: Applied Sciences 13.11 (2023), p. 6783. DOI: 10.3390/app13116783.

[8] Thomas N. Kipf and Max Welling. “Semi-Supervised Classification with Graph Convolutional Networks”. In: arXiv preprint arXiv:1609.02907 (2017).

[9] Petar Veličković et al. “Graph Attention Networks”. In: arXiv preprint arXiv:1710.10903 (2018).

[10] Ashish Vaswani et al. “Attention Is All You Need”. In: Advances in Neural Information Processing Systems 30 (NIPS 2017). 2017, pp. 5998–6008.