BÁO CÁO: TÌM KIẾM VÀ ĐÁNH GIÁ THÔNG TIN HỌC THUẬT

Ngành học: Khoa học máy tính **Chủ đề:** Ứng dụng của Học máy và Học sâu trong Phát hiện Mã độc (Malware Detection)

I. Mở đầu

- **Lý do chọn chủ đề:** Sự gia tăng nhanh chóng của các phần mềm độc hại (malware) đòi hỏi những phương pháp phát hiện tự động và thông minh hơn thay vì chỉ dựa trên chữ ký (signature-based) truyền thống. Trong lĩnh vực Khoa học máy tính, việc ứng dụng Học máy (ML) và Học sâu (DL) đang là một hướng nghiên cứu mũi nhọn để giải quyết bài toán an toàn thông tin này. Việc tìm kiếm tài liệu giúp nắm bắt các mô hình ML/DL phổ biến (như CNN, RNN, Random Forest) đang được sử dụng để phân loại mã độc.
- **Từ khóa tìm kiếm (Keywords):** "Malware detection", "Machine learning", "Deep learning", "Cybersecurity", "Phát hiện mã độc", "Học máy".

II. Quá trình và Phương pháp tìm kiếm

Để thu thập tài liệu, tôi đã sử dụng kết hợp các phương pháp và công cụ sau:

1. **Cơ sở dữ liệu học thuật:** Sử dụng **Google Scholar**, **IEEE Xplore**, **ScienceDirect** và **ACM Digital Library** để tìm kiếm các bài báo khoa học có hàm lượng chuyên môn cao.
2. **Toán tử tìm kiếm:** Sử dụng các toán tử Boolean để thu hẹp phạm vi. Ví dụ: `("Machine learning" OR "Deep learning") AND "Malware detection" AND "Survey"`. Lọc các kết quả trong khoảng thời gian từ năm 2017 đến nay để đảm bảo tính cập nhật.
3. **Sách chuyên khảo & Nguồn mở:** Tìm kiếm thêm sách nền tảng về Deep Learning và các báo cáo bảo mật từ các công ty công nghệ lớn (Microsoft, Kaspersky) trên Internet để có góc nhìn thực tiễn.

III. Tiêu chí đánh giá độ tin cậy của nguồn thông tin

Mỗi tài liệu được đánh giá dựa trên 5 tiêu chí:

4. **Tác giả (Authority):** Học hàm, học vị, và lịch sử nghiên cứu của tác giả trong lĩnh vực Khoa học Máy tính/An toàn thông tin.
5. **Cơ quan xuất bản (Publisher):** Mức độ uy tín của tạp chí/hội nghị (ưu tiên các tạp chí thuộc danh mục ISI/Scopus như IEEE, Elsevier, ACM) hoặc nhà xuất bản sách.
6. **Phương pháp nghiên cứu (Methodology):** Nghiên cứu có mô tả rõ tập dữ liệu (dataset), kiến trúc mô hình thuật toán và các chỉ số đánh giá độ chính xác (Accuracy, F1-Score) hay không.

7. **Trích dẫn (Citations):** Số lượng trích dẫn trên Google Scholar phản ánh mức độ ảnh hưởng của tài liệu trong cộng đồng nghiên cứu.
8. **Tính cập nhật (Currency):** Đặc thù ngành Công nghệ thông tin thay đổi liên tục, do đó các tài liệu xuất bản trong vòng 5-7 năm gần đây được đánh giá cao hơn.

IV. Bảng tổng hợp và đánh giá tài liệu

STT	Tên tài liệu & Tác giả (Năm)	Nguồn / Loại tài liệu	Đánh giá (Tác giả, CQXB, PP Nghiên cứu, Trích dẫn, Cập nhật)	Xếp hạng
			Tác giả: Chuyên gia tại ĐH Sapienza.	
1	*Survey of machine learning techniques for malware analysis* (Ucci et al., 2019)	Bài báo KH / ScienceDirect (Elsevier)	**CQXB:** Tạp chí *Computers & Security* (Q1). **PP:** Tổng quan hệ thống minh bạch. **Trích dẫn:** >1500 lượt. **Cập nhật:** Rất tốt (2019). **Tác giả:** Các GS từ Úc và Ấn Độ.	**Rất cao**
2	*Deep learning approach for intelligent intrusion detection...* (Vinayakumar et al., 2019)	Bài báo KH / IEEE Xplore	**CQXB:** *IEEE Access* (Uy tín cao). **PP:** Thực nghiệm định lượng trên nhiều dataset. **Trích dẫn:** >2000 lượt. **Cập nhật:** Tốt.	**Rất cao**
3	*The rise of machine learning for detection...* (Gibert et al., 2020)	Bài báo KH / ScienceDirect	**CQXB:** *Journal of Network and Computer Applications* (Q1). **PP:** Phân tích định tính & xu hướng. **Trích dẫn:** >900 lượt. **Cập nhật:** Rất mới (2020).	**Rất cao**
4	*A survey on malware detection using data mining techniques* (Ye et al., 2017)	Bài báo KH / ACM Digital Library	**CQXB:** *ACM Computing Surveys* (Tạp chí hàng đầu ngành CS). **Trích dẫn:** >1200 lượt. **Cập nhật:** Hơi cũ nhưng mang tính nền tảng. **Tác giả:** GS ĐH California.	**Cao**
5	*Malware images: visualization and automatic classification* (Nataraj et al., 2011)	Bài báo Hội nghị / ACM	**PP:** Đề xuất phương pháp chuyển mã độc thành ảnh để dùng ML phân loại (Đột phá).	**Cao**

			Trích dẫn: >2500 lượt. **Cập nhật:** Cũ, nhưng là bài báo nền tảng bắt buộc phải đọc. **Tác giả:** Giảng viên CNTT trong nước.	
6	*Đánh giá các mô hình học sâu trong phát hiện mã độc...* (Nguyen & Pham, 2022)	Bài báo KH / Tạp chí KHTTTT (VN)	**CQXB:** Tạp chí cấp quốc gia. **PP:** Thực nghiệm thực tế trên Android. **Trích dẫn:** Vài lượt. **Cập nhật:** Rất mới. **Tác giả:** Ian Goodfellow (Cha đẻ mạng GAN).	**Trung bình - Cao**
7	*Deep learning* (Goodfellow et al., 2016)	Sách chuyên khảo / MIT Press	**CQXB:** MIT Press (Cực kỳ uy tín). **PP:** Lý thuyết thuật toán nền tảng. **Trích dẫn:** >60,000 lượt.	**Rất cao**
8	*Practical malware analysis...* (Sikorski & Honig, 2012)	Sách chuyên khảo / No Starch Press	**CQXB:** NXB uy tín về sách IT thực hành. **Cập nhật:** Đã cũ nhưng kỹ thuật phân tích vẫn áp dụng được. **Tác giả:** Tập đoàn Microsoft.	**Cao**
9	*Microsoft Digital Defense Report 2024* (Microsoft, 2024)	Báo cáo ngành / Nguồn mở Internet	**PP:** Phân tích dữ liệu thực tế từ hệ thống toàn cầu. **Trích dẫn:** Không đánh giá bằng Scholar. **Cập nhật:** Mới nhất (2024). **Tác giả:** Tác giả tự do (Freelancer).	**Cao**
10	*How AI is changing malware detection* (TechBlogX, 2025)	Bài viết Blog / Nguồn mở Internet	**CQXB:** Trang blog cá nhân không qua bình duyệt. **PP:** Nhận định chủ quan. **Cập nhật:** Rất mới (2025) nhưng độ tin cậy học thuật thấp.	**Thấp**

Xuất sang Trang tính

V. Danh mục tài liệu tham khảo (Định dạng Harvard)

9. Gibert, D., Mateu, C. and Planes, J., (2020). 'The rise of machine learning for detection and classification of malware: Research developments, trends and challenges'. *Journal of Network and Computer Applications*, 153, p.102526.

10. Goodfellow, I., Bengio, Y. and Courville, A., (2016). *Deep learning*. Cambridge, MA: MIT press.
11. Microsoft, (2024). *Microsoft Digital Defense Report 2024*. [online] Có tại: https://www.microsoft.com/en-us/security/security-insider/microsoft-digital-defense-report-2024 [Truy cập ngày 16 tháng 3 năm 2026].
12. Nataraj, L., Karthikeyan, S., Jacob, G. and Manjunath, B.S., (2011). 'Malware images: visualization and automatic classification'. In *Proceedings of the 8th international symposium on visualization for cyber security* (pp. 1-7). ACM.
13. Nguyen, T.A. and Pham, V.H., (2022). 'Đánh giá các mô hình học sâu trong phát hiện mã độc trên nền tảng Android'. *Tạp chí Khoa học Công nghệ Thông tin và Truyền thông*, 12(1), pp.20-30.
14. Sikorski, M. and Honig, A., (2012). *Practical malware analysis: the hands-on guide to dissecting malicious software*. San Francisco: No Starch Press.
15. TechBlogX, (2025). *How AI is changing malware detection*. [online] (Nguồn giả định dùng để minh họa nguồn có độ tin cậy thấp).
16. Ucci, D., Aniello, L. and Baldoni, R., (2019). 'Survey of machine learning techniques for malware analysis'. *Computers & Security*, 81, pp.123-147.
17. Vinayakumar, R., Alazab, M., Soman, K.P., Poornachandran, P., Al-Nemrat, A. and Venkatraman, S., (2019). 'Deep learning approach for intelligent intrusion detection system'. *IEEE Access*, 7, pp.41525-41550.
18. Ye, Y., Li, T., Adjeroh, D. and Iyengar, S.S., (2017). 'A survey on malware detection using data mining techniques'. *ACM Computing Surveys (CSUR)*, 50(3), pp.1-40.