

**A systematic literature review on Windows malware detection: Techniques,
research issues, and future directions**

Critical Review Paper (CRP)



Anggota Kelompok :

Adya Handika Putra AP	242410101014
Talitha Puspita Sari	242410101017
Yosico Fathaura Almeivano	242410101059
Djonathan Paulsen Manik Raja J	242410101087

SEMESTER GENAP
FAKULTAS ILMU KOMPUTER
UNIVERSITAS JEMBER
TAHUN 2026

ABSTRAK

Makalah ini bertujuan untuk melakukan tinjauan kritis terhadap artikel Systematic Literature Review (SLR) berjudul “A Systematic Literature Review on Windows Malware Detection Techniques”. Artikel tersebut membahas teknik deteksi malware yang menjadi gangguan pada sistem operasi Windows dengan pendekatan machine learning dan deep learning. Review ini dilakukan melalui analisis terhadap metodologi penelitian, hasil sintesis, serta kontribusi ilmiah serta gap penelitian yang diidentifikasi dalam konteks cybersecurity yang diberikan oleh artikel tersebut.

Temuan utama review menunjukkan bahwa artikel menggunakan metodologi sistematis berbasis protokol Kitchenham, menyajikan sintesis komprehensif terkait teknik deteksi, dataset benchmark, metrik evaluasi, serta bias eksperimental seperti temporal bias dan spatial bias. Namun demikian, terdapat keterbatasan dalam pembahasan implementasi praktis, pembaruan data ancaman terkini, serta generalisasi model lintas dataset.

Disimpulkan bahwa studi ini memberikan kontribusi penting untuk pengembangan dan pengambilan keputusan di bidang deteksi malware berbasis machine learning dan deep learning dengan rekomendasi penelitian lanjutan untuk mengatasi bias, pengembangan dataset representatif, dan eksplorasi isu keamanan pembelajaran mesin khususnya dalam konteks professional issues seperti keamanan data, etika siber, dan tata kelola keamanan informasi.

Kata kunci : Malware Windows, systematic literature review, machine learning, cybersecurity ethics, professional issues

PENDAHULUAN

1.1 Latar Belakang

Perkembangan malware dalam satu dekade terakhir menunjukkan peningkatan kompleksitas yang signifikan. Sistem operasi Windows, yang menguasai lebih dari 70% pangsa pasar desktop global, menjadi target utama serangan siber karena prevalensinya yang masif dalam lingkungan individu maupun korporasi. Ransomware modern tidak hanya mengenkripsi data korban, tetapi juga melakukan pengambilan data sebelum meminta tebusan. Sistem operasi Windows menjadi target utama serangan malware global karena dominasi penggunaannya dalam lingkungan individu maupun organisasi (Maniriho et al., 2024).

Dalam konteks Professional Issues, keamanan sistem Windows berkaitan langsung dengan, Pertama etika profesi TI, khususnya tanggung jawab moral para insinyur perangkat lunak dalam merancang sistem yang aman dan dapat diaudit. Kedua tata kelola keamanan informasi (IT Governance), yang mencakup kebijakan, prosedur, dan akuntabilitas dalam manajemen risiko siber, Ketiga keberlanjutan organisasi (sustainability), karena kegagalan sistem keamanan dapat menghancurkan kelangsungan operasional bisnis, dan keempat perlindungan data dan privasi, yang semakin diatur secara ketat oleh regulasi di Uni Eropa dan berbagai legislasi nasional lainnya.

Kegagalan dalam mendeteksi malware tidak hanya berdampak teknis, tetapi juga berdampak pada reputasi organisasi, kerugian finansial, serta pelanggaran etika profesional. Oleh karena itu, keberadaan Systematic Literature Review (SLR) yang memetakan teknik deteksi malware menjadi sangat relevan. Dalam bidang rekayasa perangkat lunak, Kitchenham dan Brereton telah lama menetapkan panduan SLR yang menjadi standar komunitas riset. secara khusus mendokumentasikan pelajaran penting dari penerapan SLR dalam domain rekayasa perangkat lunak (Brereton et al., 2007).

Artikel yang direview berupaya menyajikan pemetaan komprehensif terhadap teknik deteksi malware Windows berbasis machine learning dan deep learning. Evaluasi kritis terhadap artikel ini penting untuk menilai kualitas metodologinya serta relevansinya terhadap isu profesional.

1.2 Tujuan Review

Tujuan dari Critical Review Paper ini adalah:

1. Menganalisis kualitas metodologi SLR yang digunakan.
2. Mengevaluasi kontribusi ilmiah artikel terhadap bidang deteksi malware.
3. Mengidentifikasi kelemahan dan gap penelitian.
4. Menilai relevansi artikel terhadap Professional Issues dalam cybersecurity.

RINGKASAN ARTIKEL SLR YANG DIREVIEW

Judul Artikel : A Systematic Literature Review on Windows Malware Detection Techniques

Penulis : Pascal Maniriho, Abdun Naser Mahmood, Mohammad Javed Morshed Chowdhury
Department of Computer Science and Information Technology, La Trobe University, Melbourne,
VIC, Australia

Tahun : March 2024

Jurnal Proceeding : Journal of Systems and Software, Volume 209, Article 111921, Elsevier.
DOI: <https://doi.org/10.1016/j.jss.2023.111921>

Tujuan SLR

Menyajikan studi komprehensif mengenai teknik deteksi malware pada sistem operasi Windows, termasuk klasifikasi metode, dataset, metrik evaluasi, dan tantangan penelitian. Tujuan utama dari penelitian ini adalah untuk meninjau studi mengenai deteksi malware pada sistem operasi Windows. Penelitian ini mensintesis pengetahuan dari berbagai studi tersebut untuk memahami dengan lebih baik berbagai topik terkait deteksi malware serta mengidentifikasi isu-isu penelitian baru yang berpotensi menjadi arah penelitian di masa depan. Secara lebih spesifik, studi Systematic Literature Review (SLR) ini bertujuan untuk menjawab pertanyaan-pertanyaan penelitian.

Research Questions (RQ)

Secara umum, ada 7 RQ dalam artikel mencakup:

1. Teknik apa saja yang digunakan dalam deteksi malware Windows?
2. Algoritma ML/DL apa yang paling banyak digunakan?
3. Dataset dan metrik evaluasi apa yang umum digunakan?
4. Apa saja keterbatasan dan bias dalam penelitian sebelumnya?
 - RQ1 : What are the types of malware detection techniques and their deployment methods?
 - RQ2: Which public benchmark datasets and features are used for malware detection in windows desktop devices?
 - RQ3: Which ML algorithms are mostly used for detecting malware attacks?
 - RQ4: What current DL algorithms are employed to implement malware detection techniques?
 - RQ5: What are the existing evaluation metrics for assessing the performance of malware detection techniques?

- RQ6: What are the critical experimental factors/biases in ML and DL-based techniques for detecting malware?
- RQ7: What are the research challenges in windows malware analysis and detection?

Database yang Digunakan

Pencarian literatur dilakukan pada enam database ilmiah internasional, yaitu: IEEE Xplore, ACM Digital Library, ScienceDirect, Springer Link, Web of Knowledge, dan Google Scholar. Rentang tahun pencarian mencakup publikasi dari 2009 hingga 2022, memberikan cakupan historis yang memadai untuk mengamati evolusi teknik deteksi malware selama 13 tahun. Metode SLR yang diadopsi merujuk pada protokol sebagaimana didokumentasikan dalam topik jurnal (Kitchenham et al., 2009).

Jumlah Artikel Terpilih

Artikel ini melakukan proses seleksi literatur secara sistematis dan menghasilkan sejumlah studi terpilih yang memenuhi kriteria inklusi serta quality assessment. Tahap pertama adalah penentuan sumber data dan strategi pencarian (defining data sources and search strategy), yang menghasilkan sebanyak 418 makalah. Pada tahap ini ditentukan basis data yang digunakan, seperti IEEE Xplore, ACM Digital Library, ScienceDirect, Springer, Web of Science, dan Google Scholar. Selain itu, disusun pula search query atau kata kunci menggunakan Boolean operator, yaitu : (“machine learning” OR “deep learning” OR “Windows” OR “static analysis” OR “behaviour analysis” OR “hybrid analysis”) AND (“malware detection” OR “malware classification”). Proses seleksi dilakukan secara bertahap. Pencarian awal menghasilkan 418 makalah, kemudian setelah penerapan kriteria inklusi dan eksklusi jumlahnya berkurang menjadi 320 makalah, dan setelah dilakukan quality assessment, jumlah studi primer yang dianalisis menjadi 219 makalah.

Kriteria inklusi yang digunakan meliputi :

1. Istilah pencarian muncul dalam judul, abstrak, atau kata kunci.
2. Makalah memperkenalkan teknik deteksi malware.
3. Penelitian menggunakan algoritma Machine Learning atau Deep Learning.
4. Makalah menyajikan analisis eksperimental.

Sementara itu, kriteria eksklusi mencakup makalah yang merupakan duplikasi, tidak menggunakan bahasa Inggris, teks lengkap tidak tersedia, berupa poster atau catatan pendek, serta studi yang tidak berkaitan dengan deteksi malware pada sistem operasi Windows.

Metode Seleksi

Menggunakan protokol SLR berdasarkan Kitchenham et al. (2009), dengan tahapan:

1. Perumusan RQ
2. Penentuan kriteria inklusi & eksklusi
3. Proses penyaringan bertahap
4. Quality assessment

Hasil Utama

1. Mayoritas penelitian menggunakan ML dan DL.
2. CNN dan LSTM menjadi algoritma populer.
3. Akurasi deteksi banyak dilaporkan di atas 90%.
4. Terdapat bias eksperimental seperti temporal bias dan sample imbalance.

Kesimpulan Penulis

Penulis menyimpulkan bahwa pendekatan ML/DL menjanjikan dalam deteksi malware Windows, namun masih terdapat tantangan substansial terkait bias eksperimen, generalisasi model lintas dataset, deteksi zero-day malware, dan kebutuhan akan model yang lebih explainable. Arah penelitian masa depan yang direkomendasikan meliputi federated learning untuk preservasi privasi, explainable AI (XAI), dan deteksi real-time

ANALISIS KRITIS

3.1 Analisis Metodologi

Merumuskan tujuh RQ yang mencakup spektrum luas dari teknik deteksi, dataset, algoritma ML dan DL, metrik evaluasi, bias eksperimental, hingga tantangan penelitian. Secara umum, RQ-RQ tersebut jelas dan terukur, terutama RQ2 ("Dataset benchmark publik dan fitur apa yang digunakan?"), RQ3 ("Algoritma ML mana yang paling banyak digunakan?"), dan RQ5 ("Apa metrik evaluasi yang ada?"). Pertanyaan-pertanyaan ini menghasilkan jawaban kuantitatif yang dapat diverifikasi melalui tabel dan grafik (Maniriho et al., 2024) .

Namun, RQ1 dan RQ7 bersifat lebih kualitatif dan luas. RQ7 ("Apa saja tantangan penelitian dalam analisis dan deteksi malware Windows?") mencakup berbagai isu heterogen seperti adversarial attacks, concept drift, dan kurangnya teknik hibrida. Ketidakspesifikan RQ7 berpotensi membuat sintesis menjadi deskriptif daripada analitis mendalam. Dari perspektif akademik, keterukuran RQ yang ideal mensyaratkan operasionalisasi yang dapat menghasilkan jawaban konkret dan dapat direplikasi (Kitchenham et al., 2009).

Strategi pencarian menggunakan ekspresi Boolean dengan kombinasi istilah kunci seperti "machine learning" OR "deep learning" OR "Windows" AND "malware detection" OR "malware classification". Pendekatan ini memenuhi standar SLR karena transparan dan dapat direplikasi. Penggunaan enam database bereputasi tinggi memperluas cakupan literatur secara signifikan.

Namun, terdapat potensi kelemahan: istilah "Windows" tanpa konteks spesifik dapat menghasilkan noise dalam pencarian, dan tidak ada penjelasan mengenai bagaimana konflik antar penulis dalam seleksi makalah diselesaikan. Praktik terbaik SLR mengharuskan adanya mekanisme resolusi disagreement antar reviewer. Artikel juga tidak menyertakan PRISMA flow diagram yang kini menjadi standar dalam pelaporan SLR

Pemilihan IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, Web of Science, dan Google Scholar merupakan kombinasi yang sangat relevan untuk domain keamanan siber dan rekayasa perangkat lunak. Tambahan konferensi keamanan komputer internasional terkemuka memperkuat cakupan. Database ini konsisten dengan SLR serupa di bidang yang sama.

Artikel menyajikan lima kriteria inklusi dan enam kriteria eksklusi yang relatif jelas. Pembatasan pada file executable Windows (.EXE) memberikan fokus yang tepat dan membedakan studi ini dari tinjauan malware mobile atau IoT. Batasan bahasa (hanya bahasa Inggris) merupakan praktik umum namun dapat menimbulkan language bias yang mengecualikan penelitian relevan dalam bahasa lain.

Artikel tidak menjelaskan apakah proses seleksi makalah dilakukan secara independen oleh minimal dua reviewer, yang merupakan persyaratan metodologis standar untuk meminimalkan selection bias. Pernyataan bahwa "langkah-langkah dilakukan melalui kolaborasi semua penulis" tidak cukup spesifik untuk memastikan independensi penilaian. Selain itu, keterbatasan pada publikasi berbahasa Inggris dan periode 2009 - 2022 mungkin mengecualikan studi relevan di luar rentang tersebut.

Quality assessment menggunakan delapan kriteria dengan skala tiga poin (skor minimum 4 dari 8) merupakan kontribusi positif yang mengikuti panduan Kitchenham et al. (2009). Namun, reliabilitas antar-penilai (inter-rater reliability) tidak dilaporkan. Tanpa koefisien Cohen's Kappa atau metrik serupa,

konsistensi penilaian tidak dapat diverifikasi secara independen, yang melemahkan kredibilitas quality assessment. Ini merupakan gap metodologis yang signifikan.

3.2 Analisis Hasil dan Sintesis

Sintesis dalam artikel ini dilakukan terutama melalui narrative synthesis yang dikombinasikan dengan kategorisasi kuantitatif berupa frekuensi penggunaan teknik, algoritma, dan fitur. Pengelompokan teknik deteksi ke dalam tiga kategori statis, dinamis, dan hibrida yang merupakan taksonomi yang sudah dikenal dan konsisten dengan literatur sebelumnya

Kekuatan sintesis terletak pada cakupannya yang luas dan penyajiannya yang terstruktur. Namun, sintesis tersebut cenderung bersifat deskriptif-enumeratif daripada analitik-interpretatif. Misalnya, artikel mencatat bahwa CNN dan LSTM mendominasi pendekatan DL, tetapi tidak menganalisis secara mendalam mengapa arsitektur ini lebih disukai dibanding alternatif seperti Transformer atau Graph Neural Network (GNN), atau bagaimana tren ini mencerminkan evolusi paradigma representasi pengetahuan dalam keamanan siber. Analisis mendalam semacam ini justru lebih relevan bagi pembuat kebijakan dan praktisi yang perlu memahami alasan di balik rekomendasi teknis.

Artikel berhasil memetakan tren dengan menunjukkan bahwa: (1) teknik berbasis DL meningkat pesat pada 2022 (37 studi) dibandingkan teknik berbasis ML (9 studi pada periode yang sama), (2) CNN adalah algoritma DL paling populer (21 studi), dan (3) dataset berbasis urutan API calls paling banyak diproduksi. Pemetaan tren ini bermanfaat untuk memandu prioritas penelitian masa depan.

Artikel tidak hanya bersifat deskriptif, tetapi juga mengelompokkan teknik deteksi berdasarkan pendekatan analisis (static, dynamic, hybrid). Sintesis hasil cukup komprehensif, namun sebagian besar masih berupa pengelompokan tanpa meta-analisis kuantitatif. Pemetaan tren algoritma sudah dilakukan, tetapi kurang mendalam dalam menganalisis evolusi performa dari waktu ke waktu. Gap penelitian diidentifikasi, terutama terkait bias eksperimental dan kebutuhan validasi realistik. Artikel mengidentifikasi beberapa gap penting: ketiadaan teknik deteksi hibrida berbasis DL, kurangnya studi berbasis analisis memori, dan minimnya dataset hybrid (statis, dinamis, memori). Gap-gap ini diidentifikasi secara sistematis berdasarkan data, bukan sekadar opini, menjadikannya kontribusi yang dapat diverifikasi. Namun, prioritas relatif antar gap tidak dibahas, sehingga peneliti masa depan tidak mendapat panduan yang jelas tentang gap mana yang paling mendesak untuk diatasi.

3.3 Kontribusi terhadap Professional Issues

Artikel secara tidak langsung relevan dengan dimensi etika profesi dalam keamanan siber. Diskusi tentang adversarial attacks dan mekanisme penghindaran malware baru menggarisbawahi tanggung jawab etis profesional keamanan untuk terus memperbarui kapabilitas pertahanan. Penyebutan bahwa 86,2% organisasi pernah dikompromikan oleh setidaknya satu serangan siber menegaskan urgensi tanggung jawab profesional (Maniriho et al., 2024).

Akan tetapi, artikel tidak membahas secara eksplisit dimensi etika dalam penelitian keamanan, seperti tanggung jawab dalam pengungkapan kerentanan (responsible disclosure), penggunaan data malware yang diperoleh secara etis, atau implikasi dual use dari pengetahuan adversarial attacks yang dipublikasikan. Ini merupakan keterbatasan dari perspektif professional ethics.

Temuan bahwa sektor keuangan, kesehatan, ritel, dan pendidikan adalah yang paling rentan terhadap serangan malware memberikan informasi berbasis bukti yang relevan untuk perumusan kebijakan keamanan siber sektoral. Kompilasi 29 dataset benchmark yang tersedia secara publik juga bermanfaat langsung bagi praktisi yang ingin mengembangkan atau memvalidasi sistem deteksi malware mereka.

Identifikasi bias eksperimental memiliki implikasi praktis yang kuat jika teknik deteksi malware dikembangkan dengan bias eksperimental, kinerjanya di lingkungan produksi akan jauh lebih rendah dari yang dilaporkan dalam penelitian. Peringatan ini langsung relevan bagi praktisi yang mengevaluasi solusi keamanan berbasis klaim penelitian akademis.

Artikel ini turut memberikan landasan akademik yang dapat digunakan sebagai dasar dalam penyusunan kebijakan keamanan siber di lingkungan organisasi. Meskipun demikian, pembahasan mengenai implikasi praktis bagi manajemen organisasi masih belum dijelaskan secara mendalam.

3.4 Kelebihan Artikel

- Penelitian ini menggunakan metodologi yang sistematis dengan berlandaskan pada protokol yang jelas dan terstruktur.
- Studi ini memanfaatkan database yang luas serta berasal dari sumber yang kredibel sehingga mendukung keandalan data yang digunakan.
- Penelitian ini menyajikan sintesis yang komprehensif terhadap berbagai teknik machine learning dan deep learning.
- Penelitian ini secara eksplisit mengidentifikasi adanya potensi bias eksperimental yang dapat memengaruhi hasil penelitian.
- Proses seleksi dalam penelitian ini menunjukkan tingkat transparansi yang relatif baik sehingga dapat meningkatkan kepercayaan terhadap hasil yang diperoleh.

3.5 Kelemahan Artikel

- Penelitian ini masih kurang membahas implementasi sistem dalam konteks penggunaan secara real-time.

- Studi ini tidak mengkaji aspek adversarial machine learning secara mendalam sehingga analisis terhadap potensi serangan terhadap model masih terbatas.
- Pembahasan mengenai dampak praktis dari temuan penelitian terhadap kebijakan organisasi masih tergolong minim.
- Penelitian ini tidak melakukan meta analisis statistik yang dapat memperkuat sintesis hasil dari berbagai studi yang dianalisis.
- Data statistik mengenai ancaman yang digunakan dalam penelitian ini dinilai kurang mutakhir sehingga mungkin tidak sepenuhnya merepresentasikan kondisi terkini.

3.6 Research Gap dan Rekomendasi

- Beberapa gap yang teridentifikasi : Penelitian ini masih memiliki kekurangan seperti proses validasi model pada berbagai dataset yang berbeda. Penelitian terkait juga masih menunjukkan keterbatasan dalam bentuk minimnya studi longitudinal yang memanfaatkan data berbasis waktu nyata. Terdapat kebutuhan untuk mengintegrasikan pendekatan explainable AI guna meningkatkan transparansi dan interpretabilitas model yang digunakan. Penelitian di masa mendatang perlu melakukan pengujian empiris dalam lingkungan enterprise yang nyata agar hasilnya lebih relevan secara praktis. Selain itu, diperlukan pengembangan pendekatan yang lebih kuat dalam hal robustness guna menghadapi potensi adversarial attack terhadap sistem.
- Rekomendasi : Penelitian selanjutnya perlu mengembangkan studi eksperimental yang berbasis pada data real-world agar hasil yang diperoleh lebih merepresentasikan kondisi nyata. Penggunaan protokol evaluasi yang standar juga diperlukan untuk memastikan konsistensi serta keterbandingan hasil antar penelitian. Integrasi aspek etika AI dalam pengembangan model deteksi juga menjadi hal penting untuk memastikan bahwa sistem yang dihasilkan tetap memperhatikan prinsip tanggung jawab dan keadilan dalam penerapannya. Selain itu, perlu dikembangkan sebuah framework benchmarking yang bersifat terbuka guna memudahkan peneliti lain dalam melakukan evaluasi dan perbandingan model.
- Tidak ada meta-analisis kuantitatif : Sintesis hanya bersifat naratif-deskriptif. Potensi meta-analisis untuk membandingkan akurasi atau F1-score antar teknik tidak dieksploitasi, meskipun heterogenitas data mungkin menjadi kendala.
- Tidak membahas etika penelitian keamanan: Aspek responsible disclosure, penggunaan data malware yang etis, dan implikasi dual use dari adversarial attack research tidak dibahas, padahal relevan dengan professional issues di bidang keamanan siber.

KESIMPULAN

Menyajikan SLR yang kuat dan komprehensif tentang deteksi malware Windows yang melampaui tinjauan sebelumnya dalam hal jumlah makalah yang dikaji (219 dengan maksimum 184 dalam survei

terdahulu), kelengkapan katalogisasi dataset dan algoritma, serta kontribusi orisinal berupa identifikasi 10 bias eksperimental kritis.

Dari perspektif kualitas metodologi, artikel memenuhi banyak persyaratan SLR yang baik: mengikuti protokol, menggunakan database yang relevan dan luas, serta menyajikan kriteria seleksi yang transparan. Namun, terdapat beberapa kelemahan metodologis yang perlu diperhatikan, terutama ketiadaan inter-rater reliability reporting, absennya PRISMA flow diagram, dan tidak dibahasnya publication bias. Dari perspektif kedalaman analisis, sintesis cenderung deskriptif-naratif dengan keterbatasan analisis komparatif kuantitatif. Identifikasi bias eksperimental merupakan kontribusi analitis terkuat yang memberikan nilai tambah nyata bagi peneliti dan praktisi. Dari perspektif professional issues, artikel memberikan implikasi praktis yang signifikan: taksonomi teknik deployment yang bermanfaat untuk pengambilan keputusan organisasi, peringatan tentang bias eksperimental yang relevan untuk evaluasi solusi keamanan komersial, dan identifikasi gap penelitian yang di bidang keamanan siber.

Kesimpulannya, artikel ini layak dijadikan referensi utama dalam kajian deteksi malware Windows, dengan catatan bahwa pembaca perlu mewaspadaai keterbatasan metodologis yang telah diidentifikasi. Studi lanjutan yang mengatasi gap metodologis dan penelitian yang teridentifikasi akan memberikan kontribusi signifikan terhadap kemajuan ilmu keamanan siber dan perlindungan infrastruktur digital global.

DAFTAR PUSTAKA

- Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, 80(4), 571–583. <https://doi.org/10.1016/j.jss.2006.07.009>
- Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review.

Information and Software Technology, 51(1), 7–15.

<https://doi.org/10.1016/j.infsof.2008.09.009>

Maniriho, P., Mahmood, A. N., & Chowdhury, M. J. M. (2024). A systematic literature review on Windows malware detection: Techniques, research issues, and future directions. *Journal of Systems and Software*, 209, 111921. <https://doi.org/10.1016/j.jss.2023.111921>