ISSN: 2276-4201
Volume 16, Number 2, 2026
International Journal of Innovations in Engineering Science and Advanced Technology
Kwame Nkrumah' University Science and Technology, Kumasi, Ashaniti, Ghana
Journal Homepage: www.bwjournal.org

CYBER RISK ANALYSIS AND THREAT ASSESSMENTS FROM GOVERNMENT OPEN DATASETS ON EDUCATION USING HYBRID DEEP LEARNING MODEL: A LITERATURE REVIEW

Abdullahi Yushau Ekundayo
Department of Computer Science, Faculty of Computing, Abdullahi Fodiyo University of Science and Technology, Aliero.

Prof A.B Garko
Department of Computer Science, Faculty of Computing, Northwest University Kano

Dr. M. S. Argungu
Department of Computer Science, Faculty of Computing, Abdullahi Fodiyo University of Science and Technology, Aliero.

Dr. Atiku A. Muslim,
Faculty of Engineering, Abdullahi Fodiyo University of Science and Technology, Aliero.

Abstract
The institutionalization of open data policies and the digitization of educational systems have jointly produced large, publicly accessible data ecosystems that were not originally designed with cybersecurity analytics in mind. While government open datasets on education are widely used for transparency, planning, and research, their unintended role as indicators of cyber risk has received limited scholarly attention. This literature review examines how contemporary machine learning and deep learning approaches particularly hybrid deep learning models have been employed for cyber risk analysis and threat assessment, and evaluates their relevance to education-sector government open data. Drawing on peer-reviewed studies across cybersecurity, artificial intelligence, and open data governance, the review synthesizes methodological trends, model architectures, and empirical findings. The analysis demonstrates that hybrid deep learning frameworks integrating convolutional, recurrent, and attention-based components consistently outperform single-model approaches in cybersecurity contexts. However, their application to education-related open datasets remains sparse, indirect, and methodologically underdeveloped. The review identifies key technical, data-related, and governance challenges and argues for a shift toward explainable, adaptive, and policy-aligned cyber risk analytics tailored to education-sector open data environments.

Keywords: Cyber risk analysis, Government open data, Hybrid deep learning models, Education sector cybersecurity

1. Introduction

Government open data initiatives have fundamentally altered how public-sector information is produced, disseminated, and reused. In the education sector, national and subnational governments routinely publish datasets covering institutional finance, enrollment patterns, infrastructure development, staffing, and performance indicators. These datasets are primarily intended to enhance transparency and inform policy decisions. However, as education systems become increasingly digitized, the same datasets may also function as indirect signals of cyber exposure, operational dependencies, and systemic vulnerabilities.

At the same time, educational institutions have emerged as persistent targets of cyberattacks. Unlike commercial enterprises, schools and universities often operate within open digital environments characterized by heterogeneous users, decentralized governance, and constrained cybersecurity

investment. Reports from international cybersecurity agencies consistently identify education among the sectors most affected by ransomware, phishing, and data exfiltration attacks (ENISA, 2023). Despite this risk profile, cyber risk assessment frameworks in education remain fragmented and largely reactive.

Recent advances in artificial intelligence have enabled the development of data-driven approaches to cybersecurity that move beyond signature-based detection. Deep learning models, in particular, have shown strong capacity for identifying latent patterns, anomalies, and temporal dependencies in large-scale datasets (Ismail Fawaz et al., 2019). Hybrid deep learning architectures which combine multiple neural paradigms have further improved robustness and generalization in complex cyber threat environments.

Nevertheless, existing research has overwhelmingly focused on benchmark intrusion datasets or enterprise network traffic, with limited attention to government open datasets and almost none to education-specific contexts. This review addresses that gap by examining the literature on cyber risk analysis using hybrid deep learning models and evaluating its applicability to government open education data.

2. Cyber Risk in Digitized Education Systems

Cyber risk in education cannot be fully understood through purely technical lenses. Educational institutions operate at the intersection of technology, public governance, and social access. Open networks, extensive use of personal devices, and high user turnover create conditions that amplify exposure to cyber threats (Fouad, 2021).

Empirical studies indicate that cyber incidents in education often produce cascading effects, disrupting academic delivery, administrative services, and public trust. Ransomware attacks, in particular, have caused prolonged system outages and significant financial losses across universities and public school systems (ENISA, 2023). From a risk modeling perspective, these incidents illustrate that cyber risk in education is probabilistic and systemic, rather than isolated to individual assets (Wang et al., 2020).

Traditional risk assessment approaches, which rely on expert judgment or static checklists, struggle to scale in environments characterized by rapidly changing data flows and threat vectors. This limitation has motivated increasing interest in machine learning-based risk analytics.

3. Government Open Education Data as a Cyber Risk Surface

Government open datasets are rarely classified as cybersecurity assets, yet they can inadvertently contribute to cyber exposure. Education-related open data often includes granular information on infrastructure investment, digital capacity, and institutional scale. When combined across datasets, such information may enable adversaries to infer system architectures, prioritize targets, or identify periods of heightened vulnerability.

The literature on open data governance highlights the mosaic effect, whereby sensitive insights emerge from the correlation of multiple non-sensitive datasets (Green et al., 2017). In education systems, this effect is particularly salient due to the regular publication of standardized datasets across fiscal, academic, and operational dimensions. Cremer et al. (2022) observe that most open data programs lack integrated cyber risk assessment processes, reinforcing the need for analytical tools capable of identifying latent risk patterns.

These characteristics position government open education data as a novel and underexplored input for cyber risk analytics.

4. From Machine Learning to Hybrid Deep Learning in Cybersecurity

Early applications of machine learning to cybersecurity emphasized supervised classification using handcrafted features. While methods such as Support Vector Machines and Random Forests

improved detection accuracy, they exhibited limited adaptability to evolving threats and high-dimensional data (Abraham & Bindu, 2021).

Deep learning approaches addressed these limitations by enabling automatic feature learning from raw or semi-structured data. Convolutional neural networks have been used to capture spatial patterns in network traffic, while recurrent architectures such as LSTMs model temporal dependencies in attack sequences (Ismail Fawaz et al., 2019).

Hybrid deep learning models integrate these complementary strengths. Studies consistently report that CNN–LSTM and attention-enhanced architectures outperform standalone models in intrusion detection, phishing detection, and anomaly detection tasks (Tang & Mahmoud, 2021; Lin et al., 2023). Despite these advances, the transfer of such models to government open datasets particularly in education remains largely unexplored.

5. Synthesis and Identified Research Gap

Across the reviewed literature, three dominant patterns emerge:

i. Hybrid deep learning models deliver superior predictive performance in cybersecurity contexts.
ii. Empirical validation is overwhelmingly concentrated on benchmark or enterprise datasets.
iii. Government open education datasets are rarely considered as inputs for cyber risk assessment.

This disconnect represents a significant opportunity for interdisciplinary research. Applying hybrid deep learning models to education-sector open data requires not only technical adaptation but also sensitivity to governance, ethics, and interpretability.

6. Methodological Challenges in Applying Hybrid Deep Learning to Education Open Data

Despite the demonstrated effectiveness of hybrid deep learning models in cybersecurity research, several methodological challenges limit their direct application to government open datasets in the education sector. A primary constraint is the absence of explicit cybersecurity labels within most education-focused open data repositories. Unlike benchmark intrusion detection datasets, government education datasets are typically designed for administrative transparency rather than security analytics, resulting in sparse or indirect indicators of cyber events (Cremer et al., 2022).

Another critical challenge concerns data heterogeneity. Education open datasets often combine numerical indicators, categorical descriptors, temporal records, and unstructured metadata. Hybrid deep learning models are theoretically well suited to such complexity; however, effective implementation requires careful architectural design, preprocessing strategies, and feature alignment, which are rarely addressed in existing studies (Ismail Fawaz et al., 2019).

Additionally, the dynamic nature of education systems introduces concept drift, whereby underlying data distributions evolve due to policy reforms, enrollment fluctuations, or infrastructural changes. Static deep learning models trained on historical data may therefore experience rapid performance degradation if adaptive learning mechanisms are not incorporated (Wang et al., 2020).

Finally, computational and institutional constraints must be considered. Public-sector education institutions may lack the computational resources, technical expertise, or regulatory flexibility required to deploy complex hybrid models at scale, reinforcing the gap between theoretical research and practical adoption.

7. Interpretability, Ethics, and Governance Considerations

A growing body of literature emphasizes that predictive accuracy alone is insufficient for cybersecurity applications in public-sector domains. In education, cyber risk assessments often inform administrative decisions, budget allocations, and policy interventions, necessitating a high degree of model transparency and explainability (Lin et al., 2023).

Deep learning models, particularly hybrid architectures, are frequently criticized for their "black-box" nature. Without interpretability mechanisms, stakeholders may struggle to understand why certain institutions, regions, or systems are classified as high-risk. This opacity poses challenges for accountability and trust, especially in publicly governed education systems.
Ethical considerations are also central when working with government open data. Although datasets may be legally open, their secondary use for cybersecurity analytics can raise concerns related to stigmatization, institutional profiling, or unintended disclosure of vulnerabilities. The mosaic effect further complicates ethical governance, as risk signals may emerge only through cross-dataset integration (Green et al., 2017).
These issues underscore the importance of integrating explainable AI (XAI) techniques and governance-aware design principles into hybrid deep learning frameworks intended for education-sector cyber risk analysis.

8. Emerging Research Directions

Based on the synthesis of existing literature, several priority research directions can be identified.
First, there is a need for hybrid model architectures explicitly designed for education open data, rather than adaptations of models optimized for network traffic or enterprise logs. Such architectures should accommodate mixed data types and policy-driven temporal structures.

Second, transfer learning and weak supervision represent promising strategies for overcoming label scarcity. Models pretrained on benchmark cybersecurity datasets may be fine-tuned using proxy indicators derived from education open data, such as system outages, funding anomalies, or reported service disruptions (Abraham & Bindu, 2021).

Third, future work should integrate explainability mechanisms, such as attention visualization or feature attribution, to support interpretability and policy relevance. Explainable outputs can enable education administrators and policymakers to act on model insights with greater confidence (Lin et al., 2023).
Finally, research should explore adaptive and continual learning approaches that address concept drift in evolving education systems. Such approaches are essential for maintaining model validity in dynamic institutional environments shaped by policy reforms and technological change (Wang et al., 2020).

9. Implications for Policy and Practice

The findings of this review have important implications beyond technical research. For policymakers, recognizing government open education data as a potential cyber risk surface highlights the need to integrate cybersecurity considerations into open data governance frameworks. Risk-aware data publication practices, combined with analytical monitoring, could reduce unintended exposure.
For education administrators, hybrid deep learning-based cyber risk analytics offer a pathway toward proactive risk identification, enabling earlier intervention and resource prioritization. However, successful adoption requires capacity building, cross-disciplinary collaboration, and alignment with regulatory requirements.

10. Conclusion

This literature review has examined the intersection of cyber risk analysis, government open data, and hybrid deep learning within the context of education systems. While hybrid deep learning models have demonstrated strong performance in a wide range of cybersecurity applications, their application to government open education datasets remains limited and underdeveloped.

The review highlights that education-sector open data possesses unique characteristics heterogeneity, policy-driven structure, and public accountability that distinguish it from conventional cybersecurity datasets. Addressing cyber risk in this context therefore requires more than direct methodological transfer; it demands domain-sensitive, interpretable, and governance-aware analytical frameworks.

Through synthesizing interdisciplinary research and identifying critical gaps, this review provides a foundation for future empirical studies aimed at developing robust cyber risk assessment models tailored to education-sector open data environments. Such efforts are essential for strengthening the cyber resilience of education systems in an increasingly data-driven public sector.

Table 2.1: Summary of Related Literature — Cyber Risk, Open Data, and Deep Learning

Author(s) & Year	Focus Area	Methods/Models	Key Findings	Gap Relevant to This Study
Fouad (2021)	Cyber threats in higher education	Policy and risk framework analysis	Identified unique vulnerability of open-access HEIs; called for national policy strategies	No predictive modelling; no Nigerian financial data analysis; no machine learning approach
Wang et al. (2020)	Quantitative cyber risk modelling	Bayesian Network (FAIR model)	Probabilistic risk model incorporating threat actors, attack vectors, organisational defences	Not applied to open education datasets; requires extensive expert elicitation unavailable in Nigeria
Sheehan et al. (2021)	Cyber risk quantification and data availability	Systematic literature review	Identified critical scarcity of open cyber risk datasets as primary barrier to quantification	No solution proposed for data scarcity; no deep learning applied
Naagas et al. (2018)	Campus network threat modelling	Threat-driven security architecture	Designed threat-aware network framework for HEIs; campus-specific attack tree modelling	Entirely qualitative; no integration with financial data, deep learning, or open data
Bowen et al. (2022)	K-12 cyber risk management	NIST CSF-based framework	Identified financial systems as high-risk; proposed continuous monitoring framework	Not applied to Nigerian context; no computational anomaly detection; no open government data

Author(s) & Year	Focus Area	Methods/Models	Key Findings	Gap Relevant to This Study
Oyewole et al. (2024)	Cybersecurity in Nigerian banking	Review and policy analysis	Called for AI-driven financial monitoring in Nigerian public institutions	Conceptual only; no model built or tested; focuses on banking not education
Nainna et al. (2020)	Cybersecurity in Nigerian tertiary institutions	Survey (14 institutions)	Found < 30% of Nigerian institutions have incident response plans; no automated monitoring	Survey methodology only; no computational approach; no financial data analysis
Zhang & Zhang (2021)	Evolution of OGD research	Bibliometric (main path analysis)	OGD expanding to developing countries; cybersecurity application underdeveloped	Cybersecurity dimension not explored; no Nigerian financial data; no machine learning
Cremer et al. (2022)	Cyber risk and data availability	Systematic review of 175 papers	Identified lack of open cyber risk data as primary barrier; financial data suggested as alternative	No implementation proposed; no deep learning applied to financial transaction sequences
Máchová & Lněnička (2017)	Open government data quality	Multi-criteria quality assessment	Significant quality variations across national OGD portals; usability concerns documented	No cybersecurity focus; no machine learning; did not cover Nigerian OGD
Ismail Fawaz et al. (2019)	Deep learning for time-series classification	Benchmark of 9 DL architectures, 85 datasets	CNN-based architectures achieve SOTA on time-series; kernel size 3–8 recommended for short sequences	No cybersecurity or government dataset application; no LSTM or Attention components
Abraham & Bindu (2021)	CNN for intrusion detection	Conv1D on DARPA dataset	CNN outperforms ML classifiers by 8–15% F1 on sequential network data	DARPA network traffic dataset; no financial data; no attention mechanism

Author(s) & Year	Focus Area	Methods/Models	Key Findings	Gap Relevant to This Study
Tang & Mahmoud (2021)	Phishing URL detection	CNN + Bidirectional LSTM (PhishNet)	Hybrid CNN-BiLSTM outperforms individual components; bidirectionality improves performance	Domain-specific URL data; no government open data; no attention mechanism
Fiore et al. (2019)	Credit card fraud detection	CNN-LSTM hybrid with sliding window	CNN-LSTM outperforms LSTM-only by 11.4% F1; sliding window approach validated	Credit card transaction domain; not applied to Nigerian government education data; no Attention
Malhotra et al. (2015)	LSTM-based time-series anomaly detection	LSTM autoencoder (unsupervised)	LSTM autoencoder detects anomalies in unlabelled time-series at 85–93% detection rate	Unsupervised setting; no financial data; no CNN or Attention; no Nigerian context
Qin et al. (2017)	CNN-LSTM-Attention for time-series prediction	DA-RNN (dual-stage attention)	Attention-augmented CNN-LSTM outperforms attention-free baseline on financial and env. data	Prediction task (not classification); no cybersecurity or government educational data
Lin et al. (2023)	DL in educational data mining	Review of 127 DL studies (2016–2022)	Zero studies found applying DL to financial cybersecurity in educational institutions	Gap confirmed: no study exists combining deep learning with educational financial data for cyber risk
Goldstein & Uchida (2016)	Comparative anomaly detection benchmark	19 algorithms across 10 datasets	No single best algorithm; pseudo-labeling from heuristics is valid for open-data settings	No deep learning; no financial data; no educational institution context
Cremer et al. (2022)	Cyber risk and data availability	Systematic review	Financial transactions identified as viable alternative to	No implementation; no model built; no Nigerian or African institutional context

Author(s) & Year	Focus Area	Methods/Models	Key Findings	Gap Relevant to This Study
			scarce incident data for ML	
ACFE (2022)	Occupational fraud in public sector	Global fraud survey (2,110 cases)	Education sector among most fraud-prone; median detection time 12 months; financial data as indicator	Survey-based; no computational modelling; no machine learning applied to fraud signals

References

Abraham, J. A., & Bindu, V. R. (2021). A deep learning approach for network intrusion detection systems. *Journal of Ambient Intelligence and Humanized Computing, 12*(1), 123–138. https://doi.org/10.1007/s12652-020-01709-1

Cremer, F., Sheehan, B., Fortmann, M., Kia, A. N., Mullins, M., & Wright, T. (2022). Cyber risk and cybersecurity: A systematic review of data availability. *The Geneva Papers on Risk and Insurance – Issues and Practice, 47*(4), 698–736. https://doi.org/10.1057/s41288-022-00266-6

ENISA. (2023). *Threat landscape for education*. European Union Agency for Cybersecurity.

Fouad, N. S. (2021). Cybersecurity challenges in higher education institutions: A policy perspective. *Journal of Higher Education Policy and Management, 43*(5), 512–528. https://doi.org/10.1080/1360080X.2021.1932025

Green, B., Chuang, J., & Lazer, D. (2017). The limits of open data in public governance. *Harvard Kennedy School Faculty Research Working Paper Series*.

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery, 33*(4), 917–963. https://doi.org/10.1007/s10618-019-00619-1

Lin, Y., Li, S., Wang, X., & Xu, L. (2023). Attention-based deep learning models for cybersecurity analytics: A survey. *IEEE Access, 11*, 12455–12478. https://doi.org/10.1109/ACCESS.2023.3240921

Tang, L., & Mahmoud, Q. H. (2021). A deep learning-based approach for phishing website detection. *IEEE Access, 9*, 15305–15317. https://doi.org/10.1109/ACCESS.2021.3053507

Wang, J., Neil, M., & Fenton, N. (2020). A Bayesian network approach for cybersecurity risk assessment. *Risk Analysis, 40*(10), 2137–2154. https://doi.org/10.1111/risa.13517

Zhang, C., & Zhang, L. (2021). Evolution of open government data research: A bibliometric analysis. *Government Information Quarterly, 38*(3), 101570. https://doi.org/10.1016/j.giq.2021.101570