Applications Of Linear Algebra in Machine Learning for Cybersecurity

Pandurang Babu Shelke
BSc. Computer Science
doi.org/10.64643/ijirt.208524-459

***Abstract*—Cybersecurity systems increasingly rely on machine learning to process security-related information and distinguish ordinary activity from behavior that may indicate a threat. This review argues that linear algebra is the mathematical layer that makes these learning applications possible by organizing security observations as vectors and matrices and by supporting transformations used before and during machine learning. Eigenvalues and eigenvectors are also important in dimensionality-reduction techniques such as Principal Component Analysis (PCA). Six review papers are discussed, covering intrusion detection, malware, anomalous behavior, phishing, and network attacks. The purpose is to connect the cybersecurity applications reported in those studies with the mathematical representation on which learning algorithms depend. This connection also provides a practical way to relate linear algebra concepts studied in computer science to cybersecurity applications.**

***Index Terms*—Linear Algebra, Machine Learning, Cybersecurity, Deep Learning, Vectors, Matrices, PCA, Intrusion Detection, Malware Detection**

I. INTRODUCTION

Modern computer security involves continuous observation of systems and networks, followed by analysis of the information collected from them. Machine learning methods are increasingly applied to this task because they can process large volumes of data and learn patterns associated with specific security conditions. The six sources used in this review illustrate the broad use of learning methods. Their topics include intrusion detection, malware analysis, anomaly detection, phishing, network attacks, and other computer-security problems. For instance, Mahdavifar and Ghorbani discuss deep-learning applications involving intrusion detection, malware detection, phishing or spam, and website-defacement detection [4].

The data used by these applications do not enter a learning algorithm in an unstructured form. Measurable characteristics of a security event can be written as numerical features and represented as a vector. A dataset containing many such events can then be viewed as a matrix. Once the information is expressed in this form, mathematical operations can be used to transform, combine, scale, or reduce the data. Eigenvalues and eigenvectors add another important connection because they occur in methods such as PCA. This review therefore argues that linear algebra is the framework that connects cybersecurity data with the machine-learning procedures applied to them.

II. LITERATURE REVIEW

Gökstorp, Katsikeas, and Johnson provide the broadest starting point among the selected sources in their 2026 review, "Machine Learning for Cybersecurity: A Comprehensive Literature Review" [1]. Their survey covered intrusion detection, malware, adversarial machine learning, phishing, traffic classification, software vulnerabilities, and privacy-related applications. Its breadth shows that machine learning is being considered in many cybersecurity situations. These situations may involve different types of information, but all require the information to be represented in a form that computational methods can process.

Berman, Buczak, Chavis, and Corbett examined deep-learning approaches for cybersecurity in their 2019 survey [2]. The architectures discussed are deep autoencoders, restricted Boltzmann machines, recurrent neural networks, and generative adversarial networks. The survey reports applications involving malware, spam, insider threats, network intrusions, false-data injection, and malicious domains. Different problems may require different model structures;

however, each model ultimately operates on numerical representations of the information supplied to it.
The 2020 survey by Choi and co-authors approaches the subject through eight computer-security challenges [3]. Its coverage includes security-oriented program analysis, return-oriented programming defence, control-flow integrity, network attacks, malware classification, system-event anomaly detection, memory forensics, and security fuzzing. Accuracy, precision, recall, and F1 score are among the measures discussed for evaluating security-oriented learning systems. This range demonstrates that learning methods can be used to study both system-level and network-level security behavior.

Mahdavifar and Ghorbani concentrate on deep learning in four cybersecurity application groups: intrusion detection, malware detection, phishing or spam detection, and website-defacement detection [4]. Their survey also considers analysis, feature extraction, preprocessing, and classification. These activities provide a direct link to linear algebra. Before a model can work with a security dataset, information must be converted into numerical features and arranged appropriately. Vector and matrix representations provide a mathematical form for this preparation and for subsequent computations.

Alghamdi's (2020) survey combines machine-learning and deep-learning methods and considers their applications, mechanisms, strengths, limitations, and future possibilities [5]. The comparison shows that no single learning technique is appropriate for every cybersecurity problem. The nature of the data and the task influence the choice of method. Regardless of the choice, the information presented to a learning system must be represented numerically. This common requirement is where vectors, matrices, and transformations become relevant.

Ali and colleagues give particular attention to intrusion-detection systems in their 2024 survey [6]. The work considers machine-learning and deep-learning strategies along with cybersecurity datasets and practical issues connected with intrusion-detection research. Intrusion records commonly contain several measurable characteristics. When these characteristics are treated as features, an individual record can be represented as a vector and a collection of records as a matrix. This provides a direct example of how a linear-algebra representation can form the starting point for machine-learning calculations in cybersecurity.

III. OVERVIEW

The reviewed studies primarily explain how machine-learning approaches are being used to address cybersecurity problems. A complementary mathematical view is obtained by considering how the underlying security data are structured. Measurements obtained from network traffic or system activity can be converted into numerical features. One security event can therefore be represented by a vector, while a collection of events can be organized as a matrix. These structures allow the data to undergo mathematical operations before entering a learning model. The resulting representation can help a model distinguish patterns associated with different security conditions.

Principal Component Analysis (PCA) provides a clear example of the role of linear algebra in this process. Cybersecurity datasets may contain many measured features, although not every feature contributes equally to the useful variation in the data. PCA constructs new directions using eigenvectors, while the associated eigenvalues indicate how much variation is captured along those directions. The original dataset can consequently be expressed through fewer components. This is a dimensionality-reduction step; it prepares the data but does not by itself perform the final cyber-threat classification or detection.

A cybersecurity machine-learning workflow can therefore be understood through its sequence of data representations and transformations. Raw security observations are first converted into numerical features and arranged in vector or matrix form. Appropriate mathematical operations can then be applied before the resulting data are supplied to a learning algorithm.
Depending on the problem, the model may be used for intrusion detection, anomaly recognition, classification, or related security tasks. Further investigation could also consider other linear-algebra approaches, including singular value decomposition and measures based on vector similarity.

IV. CONCLUSION

The six reviewed studies cover a broad set of cybersecurity applications for machine learning and deep learning, including intrusion detection, malware analysis, anomaly identification, phishing-related detection, and network security [1]– [6]. Their main emphasis is on applications, learning methods, datasets, and evaluation measures. Looking at the mathematical structure of the data adds another perspective to these applications.

Vectors and matrices provide practical representations for numerical security observations and collections of features. Operations on these structures support the preparation and transformation of data used by learning algorithms. Eigenvalues and eigenvectors are particularly relevant to dimensionality-reduction methods such as PCA because they help identify alternative directions in which the data can be represented. In this sense, linear algebra forms an underlying mathematical foundation connecting numerical cybersecurity observations with machine-learning models and the security-related outputs obtained from them.

REFERENCES

[1] S. Gökstorp, S. Katsikeas, and P. Johnson, "Machine Learning for Cybersecurity: A Comprehensive Literature Review," *ACM Computing Surveys*, 2026, doi: 10.1145/3796543.

[2] D. S. Berman, A. L. Buczak, J. S. Chavis, and C. L. Corbett, "A Survey of Deep Learning Methods for Cyber Security," Information, vol. 10, no. 4, Art. no. 122, 2019, doi: 10.3390/info10040122.

[3] Y.-H. Choi, P. Liu, Z. Shang, H. Wang, Z. Wang, L. Zhang, J. Zhou, and Q. Zou, "Using Deep Learning to Solve Computer Security Challenges: A Survey," Cybersecurity, vol. 3, Art. no. 15, 2020, doi: 10.1186/s42400-020-00055-5.

[4] P. Shigvan, S. Shevkari, V. Auti, and G. Kumar, "Detecting and mitigating insider threats with artificial intelligence," International Journal of Innovative Inventions in Social Science and Humanities, 2025.

[5] M. I. Alghamdi, "Survey on Applications of Deep Learning and Machine Learning Techniques for Cyber Security," International Journal of Interactive Mobile Technologies, vol. 14, no. 16, 2020, doi: 10.3991/ijim. v14i16.16953.

[6] S. Shevkari, S. V. Choudhari, and A. N. Tonpe, "Analysis of implementing green accounting (GA) in India—Need of the modern era," *International Journal of Scientific Development and Research*, vol. 10, no. 11, 2025.

[7] A. H. Ali *et al.*, "Unveiling Machine Learning Strategies and Considerations in Intrusion Detection Systems: A Comprehensive Survey," *Frontiers in Computer Science*, 2024.