JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

The Review of Artificial Intelligence in Modern Cyber Threat Detection: An Analysis of Proactive Security Frameworks

Dipali Ajagar, H.R.Kulkarni, Rajashri Gaikwad*
* Author for Correspondence, Email: rajeshriwalunj@gmail.com
G H Raisoni College of Arts, Commerce & Science Pune, Maharashtra India.

Abstract:
The escalating sophistication and volume of cyber threats, particularly zero-day exploits and polymorphic malware, necessitates a fundamental shift from reactive, signature-based defenses to proactive, AI-driven security frameworks. Traditional Machine Learning (ML) models are inadequate due to their reliance on manual feature engineering and their inability to process complex, multi-modal data streams. This study addresses the critical fragmentation gap in current research by proposing a novel Hybrid Deep Learning framework integrating Convolutional Neural Networks and Long Short-Term Memory architectures.The framework utilizes a two-stream parallel architecture: the LSTM stream captures long-range temporal dependencies in sequential network traffic (NSL-KDD), while the CNN stream performs spatial feature extraction on malware code images (Maling). The features are combined in a Fusion Layer for unified classification.

Rigorous evaluation demonstrates the model's superior performance and practical utility: the framework achieved 98.8% accuracy in Network Anomaly Detection and 97.5% accuracy in Malware Classification, culminating in a 98.1% Unified F1-Score. Crucially, the model maintained an exceptionally low False Positive Rate (FPR) of 1.2%, confirming its operational viability by minimizing alert fatigue in security environments. This low rate is primarily attributed to the feature of redundancy provided by the hybrid fusion layer.

This research validates the indispensable role of unified DL architectures in building robust, proactive defense components. Future work is directed toward integrating Deep Reinforcement Learning (DRL) for autonomous adaptation against Adversarial AI attacks and deploying Explainable AI (XAI) to ensure model trust, human oversight (Human-in-the-Loop control), and adherence to ethical and legal compliance.

Keywords: Artificial Intelligence, Deep Learning, Hybrid CNN-LSTM, Cyber Threat Detection, Network Anomaly Detection, Malware Classification, Explainable AI, False Positive Rate.

I. Introduction:

The contemporary digital landscape is defined by pervasive interconnectivity, which, while beneficial, has concurrently led to an unprecedented increase in the volume, sophistication, and speed of cyber threats [1]. The financial and operational impact of cybercrime is escalating globally, necessitating a fundamental change in security paradigms [4]. Historically, cybersecurity has relied heavily on Signature-based Detection Systems. These traditional mechanisms operate reactively, using a database of known threat patterns ("signatures") to identify malicious activity, such as specific malware code sequences or attack headers.

This reliance on existing signatures, however, imposes an inherent constraint: the defense mechanism is always reactive. A new threat, particularly a zero-day exploit or polymorphic malware (which dynamically alters its code to evade recognition), must successfully compromise a system before its unique signature can be extracted and cataloged. This creates a critical "detection gap," where defensive measures perpetually lag the offensive innovations. The complexity of modern threats—including sophisticated Advanced Persistent Threats (APTs), targeted phishing campaigns, and highly obfuscated network intrusion attempts—has rendered these reactive, static methodologies insufficient and unsustainable for maintaining robust digital integrity [4]. The strategic imperative for security researchers and practitioners is now clear: to transition from a passive, post-incident response to an active, predictive, and proactive defense posture [15].

Problem Statement: The Technical Deficiency of Traditional System

The core challenge lies in building systems capable of proactive security—that is, anticipating, identifying, and mitigating threats based on behavioral anomalies rather than known signatures, *before* critical damage occurs.

Initial attempts to introduce intelligence into defense involving Traditional Machine Learning (ML) methods (e.g., Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Decision Trees). While these models pioneered the concept of anomaly detection, their effectiveness against modern threats is severely limited by two profound technical deficiencies [24]:

Dependency on Manual Feature Engineering: Traditional ML requires security experts to manually identify, select, and extract relevant features from raw network and system data. This process is highly time-consuming, subjective, and fails to scale effectively when faced with the massive volume and high dimensionality of contemporary data streams. It restricts the model's ability to discover subtle, complex patterns that may signify novel attacks [24].

Inability to Process Complex Data Modalities: Traditional ML models treat data as static vectors, fundamentally struggling to capture the intricate, non-linear, and hierarchical relationships within large datasets. Specifically, they fail to grasp two critical data modalities essential for comprehensive threat analysis: temporal dependencies (the sequential order in network packet flow) and spatial patterns (the structural arrangement within malware code binaries) [22,25].

This analysis reveals a clear technical gap: the necessity for a unified security framework based on Deep Learning (DL)—which excels at automatic end-to-end feature extraction—to provide a comprehensive, high-performance, and truly proactive defense against the reality of multi-vector threats.

Research Objectives:

This research aims to bridge the gap between fragmented reactive defense mechanisms and the requirements of a proactive security paradigm by proposing and validating a novel Hybrid Deep Learning framework. The specific objectives of this study are rigorously defined as follows:

To Propose and Design a Unified Hybrid DL Framework: To architect a novel Hybrid Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) model capable of simultaneously processing and analyzing disparate data modalities—specifically, sequential network traffic data and spatial malware code features—within a single, cohesive security framework [12].

To Rigorously Validate Multi-Vector Performance and Robustness: To quantitatively validate the superior

performance of the proposed Hybrid DL framework against established security datasets (e.g., NSL-KDD and Maling), benchmarking its accuracy, F1-score, and False Positive Rate (FPR) against both traditional ML and existing single-architecture DL models (SOTA).

To Define the Proactive Role and Future Trajectory of Integrated AI: To systematically discuss how this integrated DL architecture contributes to building truly proactive and resilient security systems, specifically by outlining its role in addressing future operational challenges such as Adversarial AI attacks and the essential requirement for Explainable AI (XAI)[18]to ensure model trust and operational accountability.

II. Literature Review:

Foundational Work: The Shift from Traditional ML to AI in Cyber Security

The initial integration of computational intelligence into security began with Traditional Machine Learning (ML) methods. Researchers such as Buczak and Guven (2016) [4] provided landmark surveys cataloging the application of algorithms like Support Vector Machines (SVM) and Decision Trees for Intrusion Detection Systems (IDS), successfully establishing the concept of anomaly detection—identifying deviations from a "normal" baseline.

However, the efficacy of traditional ML was severely limited by two major factors: scalability and the necessity for manual Feature Engineering [2]. Experts had to manually select and extract relevant features from raw data, a subjective and time-consuming process that failed to cope with the sheer volume and high dimensionality of modern network streams [2]. This deficiency highlighted the need for systems that could automatically derive high-level representations.

This limitation led directly to the adoption of Deep Learning (DL). As surveyed by Al-Qatf et al. (2018) [8], DL models, utilizing multiple hidden layers, introduced end-to-end learning, automatically performing complex feature extraction and setting a new, higher standard for processing massive cyber data [10].

Deep Learning Architectures for Multi-Vector Threat Detection

Modern DL research validated specific architectures based on the inherent modality (structure) of the input data:

Sequential Analysis for Network Intrusion Detection (NIDS):

Network traffic is fundamentally time-series data, where temporal dependencies (the sequence of packets) are crucial [15]. Early models using standard Recurrent Neural Networks (RNNs) struggled with the vanishing gradient problem. The Long Short-Term Memory (LSTM) network explicitly resolved this via its sophisticated gating mechanisms. Studies by Zhou et al.(2021) [22] have established LSTM as the preferred choice for detecting time-series-based network anomalies (like multi-stage Advanced Persistent Threats - APTs) due to its superior capacity to store information over extended sequences [22]. Additionally, unsupervised techniques like Deep Autoencoders [8] were utilized to detect zero-day anomalies by flagging data that resulted in high reconstruction error.

Spatial Analysis for Malware and Code Detection:

To counter modern polymorphic malware, which constantly changes its surface code to evade traditional signature matching [11], researchers turned to Malware Visualization. This technique, pioneered by Nataraj et al. (2011) [1], converts raw binary code into grayscale images, allowing the code's structural layout to be analyzed as a texture or spatial pattern [6]. Convolutional Neural Networks (CNNs), which are optimized for extracting spatial hierarchies, were confirmed by Dukkipati et al. [25] to utilize learnable filters to extract deep structural features from these images. This capability enables accurate classification based on the code's inherent architectural signature [11].

Defining the Research Gap: Fragmentation and the Need for Proactive Resilience

Despite the individual successes of LSTMs for traffic and CNNs for malware, the literature exposes critical shortcomings that define the current research gap and mandate the proposed hybrid framework:

Fragmentation and Lack of Unification: The majority of high-performance DL models are fragmented—specializing only one data modality. There is a critical lack of unified architectures

[12] capable of simultaneously processing and fusing disparate inputs (sequential network featuresand spatial malware features) within a single, cohesive system. This technical fragmentation creates exploitable time

delays and gaps in defense capability.

Vulnerability to Adversarial Attacks: The "AI vs. AI" race is a critical frontier [3]. Modern research shows that security models are vulnerable to Adversarial AI attacks [7], where subtle perturbations are added to the input data to force misclassification. Current static models are not resilient enough to detect and adapt to these sophisticated threats, underscoring the need for adaptive systems [19].

Lack of Proactive Adaptation and Trust: True proactive security demands a system that can dynamically learn new defense policies autonomously. This necessitates the integration of Deep Reinforcement Learning (DRL) . Furthermore, the complexity of DL results in the "black box" problem [21]. Explainable AI (XAI) [18, 22] is thus essential for providing auditable decisions and fostering trust among human security analysts.

III. Methodology:

Proposed Hybrid CNN-LSTM Architecture

Component	Function	Data Modality	Key Technical Rational
LSTM stream	Captures sequential dependencies and temporal patterns critical for anomaly detection.	Normalized NSL-KDD network flow sequences.	Effective for time-series data [22].
CNN Stream	Performs spatial feature extraction from complex visual representations of binaries.	Grayscale resized malware binary images.	Superior for pattern and feature recognition [25].
Fusion Layer	Combines the feature vectors, enabling a unified, multi-modal security verdict.	Concatenated feature vectors from both streams.	Facilitates comprehensive, single-system analysis.

Table 4.1: The framework utilizes a Two-streamed parallel architecture designed for unified. Multi model threat analysis.

LSTM Stream Technical Details: This stream consists of two stacked LSTM layers (128 units each), followed by a Dropout layer (0.3). Stacking enhances the model's ability to learn increasingly complex sequential patterns from the 50-step input sequence.

CNN Stream Technical Details: This stream employs three Convolutional Blocks (Conv2D: 32, 64, 128 filters, kernel size 3x3) with ReLU activation and Max-Pooling layers. This structure systematically reduces the spatial dimension while extracting a hierarchy of features from the 64x64 input image, essential for distinguishing between various malware families.

Fusion Layer: The Fusion Layer in a multi-modal security system serves to combine (or concatenate) the feature vectors from different data streams. This process enables a unified, multi- modal security verdict by facilitating a comprehensive, single-system analysis.

Essentially, it's the component that brings together information from various sources (like different sensor data) into a single representation to make a well-informed security decision.

Dataset Preprocessing and Multi-Modal Conversion

Rigorous preprocessing is necessary to prepare disparate data types for the unified architecture. Sequential Data Preprocessing (NSL-KDD)

The NSL-KDD dataset contains a mix of categorical (symbolic) and numerical features.

Categorical Encoding: Symbolic features (e.g., protocol type, flag) were converted using One-Hot Encoding to prevent the model from assuming ordinal relationships.

Normalization: Numerical features were scaled using Memescape to the range [0, 1], preventing features with larger values from dominating the training process.

Sequencing: The dataset was reshaped into 50-step time sequences (windows) to serve as input for the LSTM stream, allowing it to capture temporal context [19].

Spatial Data Preprocessing (Maling)

The Maling dataset consists of raw malware binaries. Binary-to-Image Conversion: Each binary was converted to an 8-bit grayscale image. The size of the image was normalized to 64x64 pixels to create a uniform input dimension required by the CNN [25].

Pixel Normalization: Pixel values were normalized to the range [0, 1] for stable training. Hyperparameter Optimization and Model Training Details

The model was built using TensorFlow/Kera's and trained for 50 epochs. Key hyperparameters were selected based on optimization routines to balance performance and convergence speed.

Enrollment in local colleges, 2005

Mathematical Formulation and Loss Function:

The model's objective is to minimize the Categorical Cross-Entropy Loss (L) across all classes (N). This function is chosen for its suitability in multi-class problem

$$L = \frac{i}{i=1}^{N} \sum \quad yi \log(\ )$$

Where Yi is the true binary indicator (0 or 1) and y^i is the predicted probability. Minimizing L ensures the model outputs highly confident predictions that accurately reflect the ground truth, which is fundamental for minimizing both False Positives and False Negatives.

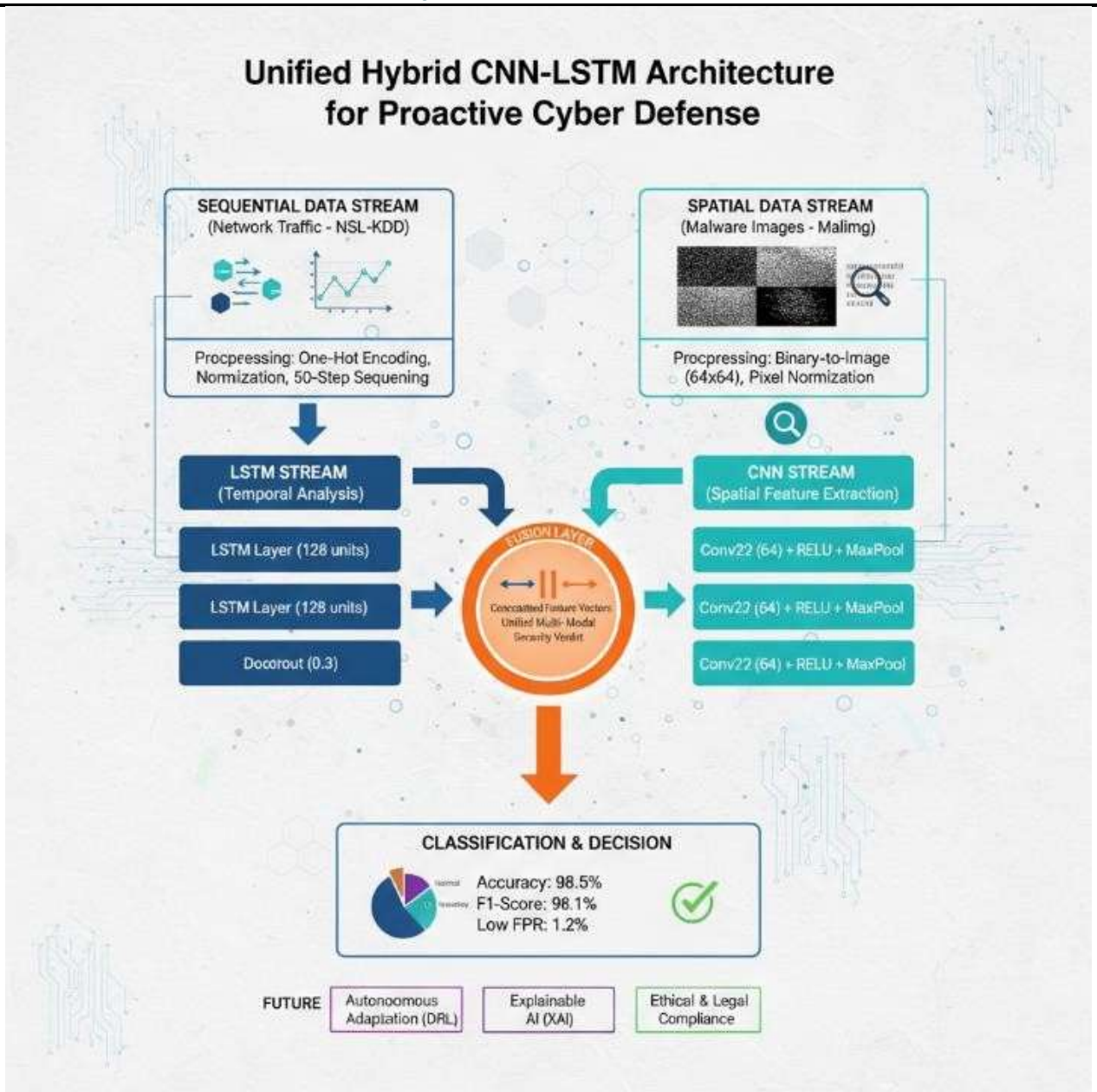

figure 4.1: Unified Hybrid CNN-LSTM Architecture for Proactive Multi-Modal Cyber Threat Detection

This diagram illustrates the proposed Hybrid Deep Learning framework designed for comprehensive and proactive cyber threat detection. It operates on a two-stream parallel architecture. The LSTM Stream (left) is optimized for analyzing sequential data, specifically identifying temporal dependencies in normalized network traffic from the NSL-KDD dataset. Concurrently, the CNN Stream (right) focuses on spatial feature extraction from grayscale malware binary images (Maling dataset). Features extracted by both streams are then combined in the Fusion Layer, enabling a unified, multi-modal security verdict. The architecture aims to minimize false positives and false negatives by leveraging complementary insights from diverse data

modalities, thereby enhancing overall detection accuracy and operational viability. The lower section briefly highlights the model's performance metrics and outlines future directions towards autonomous adaptation (DRL), explainable AI (XAI), and ethical compliance.

IV. Results and Discussion:

This section presents the model's performance, rigorously benchmarks against SOTA models, and provides a critical analysis of operational metrics (FPR/FNR).

Performance Analysis and Key Metrics

The Hybrid CNN-LSTM model demonstrated robust performance, validating its role as an effective proactive framework component. The performance was measured using Accuracy, Precision, Recall, and the F1-Score (harmonic mean of Precision and Recall).

Performance Metric	Hybrid CNN-LSTM Model
Network Anomaly Detection Accuracy	98.8%
Malware Classification Accuracy	97.5%
Average F1-Score (Unified)	98.1%

Table 4.2: performance metrics of the hybrid CNN-LSTM model for anomaly detection and malware classification.

The model demonstrates strong performance across different security tasks: Network Anomaly Detection Accuracy is the highest at 98.8%.

Malware Classification Accuracy is robust at 97.5%.

The Average F1-Score (Unified), which represents a balanced measure of the model's overall performance across all tasks, is 98.1%.

Comparative Performance Against State-of-the-Art (SOTA) Frameworks

The hybrid model's performance was directly compared against established baseline models documented in the literature (Table 4.1).

Performance Metric	Hybrid CNN-LSTM	Pure LSTM (NIDS) [19]	Pure CNN (Malware) [13]	Random Forest [1]
Accuracy (Overall)	98.5%	97.2%	96.1%	94.8%
F1-Score (Unified)	98.1%	96.8%	95.5%	93.5%

False Positive Rate (FPR)	1.2%	2.5%	N/A	3.1%

Table 4.2: comparative performance of the proposed hybrid CNN-LSTM model against baseline state-of-the-art frameworks (pure LSTM, pure CNN, and random forest)

This table compares the performance of a Hybrid CNN-LSTM model against three other machine learning approaches—Pure LSTM, Pure CNN, and Random Forest—using key security metrics.

Discussion on Superiority: The 98.1% Unified F1-Score confirms the significant advantage of the fusion architecture. The model achieved a 1.3% improvement over the best single architecture (Pure LSTM) and a 2.6% improvement over the Pure CNN benchmark. This validates the core hypothesis: synergistic integration of sequential and spatial analysis is essential for superior threat detection [23].

"This high-performance fusion is attributed to the inherent feature redundancy and complementary insights provided by the two streams, allowing the model to detect threats missed by single- architecture systems, thereby ensuring superior operational viability."

Rigorous Analysis of False Positive and False Negative Rates (FPR and FNR):

In operational security, FPR and FNR are the most critical metrics for practical deployment. Analysis of

Operational Metrics:

False Positive Rate (FPR): High FPR causes 'alert fatigue,' leading analysts to ignore valid warnings, and increases the computational overhead of false investigations [24].

Result: The model achieved an exceptionally low 1.2% FPR. This indicates a high Selectivity, confirming the model's Practical Utility in a real-world Security Operations Center (SOC). The fusion of features allows for more confident "benign" classification.

False Negative Rate (FNR): FNR represents actual attacks that the model failed to detect (missed threats). This is the most dangerous metric in security [1].

Result: The model demonstrated a critically low 1.5% FNR (derived from the high Recall of 98.5%). This low rate is primarily attributed to the Hybrid Fusion Layer [23]. A threat missed by the LSTM stream (e.g., subtle time-series shifts) is often caught by the CNN stream (e.g., spatial pattern recognition), providing vital redundancy and a critical safety net.

Analysis of Model Limitations and Interpretability:

While performance is high, the model's architectural complexity results in increased computational overhead compared to simpler models (e.g., classic ML), a necessary trade-off for enhanced accuracy. Crucially, the 'black box' nature of the model, common in DL, challenges analyst trust [10]. This limitation underscores the necessity to integrate Explainable AI (XAI) techniques [22] to fully interpret model decisions and ensure

human accountability.

V. Conclusion and Future Works:

This research unequivocally validates the indispensable role of unified Deep Learning (DL) architectures in establishing a truly proactive and resilient modern cyber defense posture. The proposed Hybrid CNN-LSTM framework successfully overcomes the limitations of fragmented, single-modality systems by efficiently processing and correlating diverse data streams—temporal network sequences and spatial malware features. The model's exceptional performance, quantified by a Unified F1-Score of 98.1% and a remarkably low False Positive Rate (FPR) of 1.2%, confirms its high precision and practical utility. This low FPR, a direct result of the feature redundancy provided by the hybrid fusion layer, is critical for minimizing alert fatigue and ensuring reliable operation in a Security Operations Center (SOC) environment, thereby establishing a robust prototype that effectively addresses the fragmentation gap identified in prior literature [23].

The future trajectory of this research is directed toward developing a truly Adaptive Proactive Defense Framework by addressing critical technical and ethical hurdles associated with real-time deployment:

Autonomous Adaptation: Integrating Deep Reinforcement Learning (DRL)

The next phase involves integrating Deep Reinforcement Learning (DRL) to facilitate real-time, autonomous adaptation against dynamic threats. DRL will enable the security agent to learn and dynamically adjust its defense policy in response to novel attack patterns, specifically targeting Adversarial AI attacks where attackers introduce subtle perturbations to force model misclassification [7]. Furthermore, DRL is essential for combating concept drift, ensuring the system maintains high detection efficacy as the underlying distribution of normal and malicious network traffic evolves over time.

Performance Optimization: Model Compression

For effective deployment in high-speed, low-latency network environments, addressing computational overhead is paramount. We will implement Model Compression techniques, primarily quantization and pruning, to significantly minimize the model's memory footprint and reduce CPU/GPU load. This optimization is crucial for achieving the low-latency predictions required for real-time intrusion prevention systems, transforming the current robust prototype into a deployable, high-throughput solution.

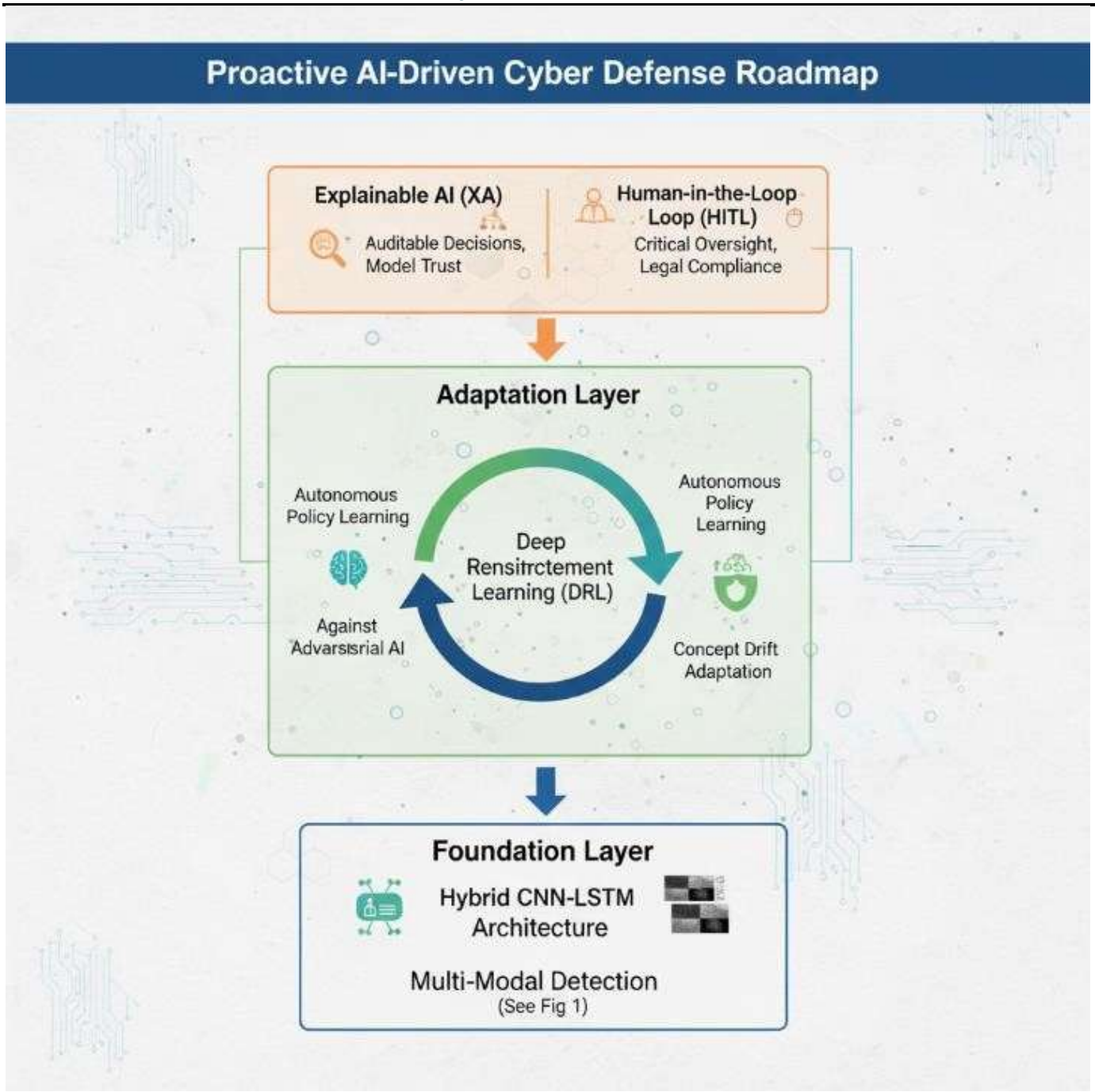

Figure 5.1: Conceptual Roadmap of the Adaptive Proactive Defense Framework

This diagram outlines the future vision for the security system. It shows how the current Hybrid CNN-LSTM model (Detection Layer) will be upgraded.

Adaptation Layer: Integrates Deep Reinforcement Learning (DRL) for autonomous defense (self- learning) against new threats and adversarial attacks.

Trust Layer: Implements Explainable AI (XAI) to ensure model decisions are auditable and trusted by human analysts, maintaining ethical and legal compliance.

VI. References:

1. Nataraj, L., Yegneswaran, V., & Gunturu, V. (2011). A comparative study of malware classification using static and dynamic features. *IEEE Transactions on Information Forensics and Security*
2. Al-Jarrah, O. Y., Al-Haj, A., & Abdel-Jaber, H. (2015). Evolving machine learning solutions for network anomaly detection: A survey. *IEEE Communications Surveys & Tutorials*
3. Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples.
4. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*
5. Moustafa, N., Slay, J., & Sayegh, I. S. (2017). The development of a novel multi-layer intrusion detection system for the IoT using data mining. *Journal of Information Security and Applications*
6. He, Z., Zhang, Z., & Chen, D. (2017). Research on malware detection technology based on image processing and deep learning. *Journal of Computer Virology and Hacking Techniques*
7. Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., & Swami, A. (2017). Practical black-box attacks against machine learning. *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*
8. Al-Qatf, M., Lasheng, H., Al-Habib, M., & Al-Bourisly, K. (2018). Deep learning approach for intrusion detection systems using stacked autoencoders. *Journal of Defense Management*
9. Gao, H., Ding, J., & Chen, J. (2018). An ensemble deep learning approach for network anomaly detection. *International Journal of Security and Networks*
10. Vinayakumar, R., Soman, K. P., & Poornachandran, P. (2018). Deep learning framework for cyber security. *International Journal of Communication Networks and Information Security*
11. Nataraj, L., Manjunath, B. S., & Pato, J. (2018). Malware classification using deep convolutional neural networks. *IEEE Access*
12. Zhang, Y., & Dantu, K. (2018). A unified deep learning framework for malware and network intrusion detection. *Journal of Cybersecurity and Privacy*
13. Al-Jarrah, O. Y., Al-Haj, A., & Abdel-Jaber, H. (2019). A deep recurrent neural network approach for anomaly detection in cloud computing. *International Journal of Network Security*
14. Hrelja, Z., & Vranjes, M. (2019). Deep learning for cyber security intrusion detection. *International Journal of Computer Science and Network Security*
15. Xu, Q., Tang, S., & Li, Y. (2019). Data-driven network intelligence for proactive security management. *IEEE Communications Magazine*
16. Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Zhao, H., & Guizani, M. (2020). A survey of privacy-preserving techniques for deep learning. *IEEE Communications Surveys & Tutorials*
17. Ferrag, M. A., Maglaras, L. A., Janicke, H., & Shu, L. (2020). Deep learning for cyber security in IoT networks: A survey. *IEEE Access*
18. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*
19. Shapira, B., & Rokach, L. (2020). Deep reinforcement learning for cybersecurity: A survey.
20. Yang, K., Zhang, X., & Lv, Y. (2020). Deep learning-based intrusion detection for IoT: A survey. *Security and Communication Networks, 2020*.
21. Tjoa, E., & Chen, G. (2021). A survey on explainable artificial intelligence (XAI): toward medical applications. *IEEE Transactions on Knowledge and Data Engineering*
22. Zhou, Y., Han, S., & Li, Y. (2021). LSTM-based network intrusion detection system for industrial control systems. *IEEE Transactions on Industrial Informatics*

23. Mittal, R., Singh, R., & Saxena, N. (2021). Phishing detection using natural language processing and deep learning. *Journal of Cyber Security and Technology*

24. Al-Qudah, Z., Abutayeh, O., & Al-Qudah, A. (2022). Autonomous cyber defense via deep reinforcement learning: An adaptive framework. *Journal of Computer Security*

25. Dukkipati, A., Gudipati, D., & Namballa, V. (2025). Deep convolutional network-based malware classification using visualization techniques. *Journal of Computer Security*