Plagiarism Checker X - Report

Originality Assessment

1%

Overall Similarity

Date: Feb 14, 2026 (12:49 PM)
Matches: 55 / 4541 words
Sources: 2

Remarks: Low similarity detected, consider making necessary changes if needed.

Verify Report:
Scan this QR Code

EXPLORING AND COMPARING THE APPLICATION OF AI 1 TRANSFORMER TECHNIQUES AND LONG-SHORT TERM MEMORY 2 IN NETWORK INTRUSION DETECTION SYSTEMS 3 4 5

Abstract 6 The increasing complexity and frequency of cyber-attacks have made Network Intrusion 7 Detection Systems (NIDS) a critical component of modern cybersecurity. Traditional machine 8 learning approaches have shown promise in detecting anomalies, but they often struggle with 9 capturing long-term dependencies and complex patterns in network traffic. This study explores 10 and compares the effectiveness of two advanced artificial intelligence techniques, Long Short11 Term Memory (LSTM) networksand Transformer-based models in the context of NIDS. LSTMs, a 12 type of recurrent neural network, are designed to handle sequential data and retain temporal 13 dependencies, making them suitable for identifying patterns over time. Transformers, leveraging 14 self-attention mechanisms, excel at modeling global relationships in sequences, enabling them to 15 capture intricate dependencies and interactions across network traffic features. By evaluating 16 both techniques on benchmark intrusion detection datasets, this study highlights differences in 17 detection accuracy, computational efficiency, and adaptability to evolving attack patterns. The 18 findings suggest that while LSTMs provide robust temporal analysis, Transformers demonstrate 19 superior performance in recognizing complex and context-dependent intrusion patterns, offering 20 a promising direction for next-generation NIDS. 21 Keyword: Long-short term Memory (LSTM), Network Intrusion Detection Systems(NIDS), AI 22 Transformer, Recurrent Neural Network(RNN), Distributed Denial-of-service (DDoS), Artificial 23 Intelligence (AI) 24

1 INTRODUCTION 25

1.1 Background review 26

The rise of digital transformation and the widespread adoption of internet-connected systems 27 have exponentially increased the volume and complexity of network traffic. Consequently, 28 organizations face a growing risk of cyber threats such as malware, ransomware, phishing, and 29 distributed denial-of-service (DDoS) attacks. Network Intrusion Detection Systems (NIDS) serve 30 as a critical defense mechanism, monitoring network traffic to identify and mitigate potential 31 security breaches. Traditional NIDS rely on signature-

based or statistical anomaly detection 32 methods, which, while effective for known attacks, struggle with detecting novel or sophisticated 33 threats. 34 Artificial Intelligence (AI) has emerged as a powerful tool for enhancing NIDS capabilities. In 35 particular, deep learning techniques have shown promise in capturing complex patterns in 36 network traffic data. Two advanced approaches, Long Short-Term Memory (LSTM) networks 37

and Transformer-based modelshave gained attention for their ability to process sequential and 38 high-dimensional data. 39 □ LSTM Networks: LSTMs are a type of recurrent neural network (RNN) designed to 40 handle sequences with long-term dependencies. They utilize memory cells and gating 41 mechanisms to retain relevant information over extended periods, making them 42 particularly effective for time-series data such as network traffic logs. LSTMs have been 43 successfully applied in anomaly detection, identifying unusual patterns indicative of 44 intrusions. However, LSTMs can be computationally intensive and may struggle to 45 capture complex global dependencies in very large datasets. 46 □ Transformer Models: Transformers, originally developed for natural language 47 processing, leverage self-attention mechanisms to capture both local and global 48 relationships in sequential data. Unlike LSTMs, Transformers process entire sequences 49 simultaneously, enabling more efficient learning of intricate dependencies and 50 interactions. In the context of NIDS, Transformers can analyze large volumes of network 51 traffic with high contextual awareness, potentially improving detection of sophisticated, 52 context-dependent attacks. 53 Comparing these two approaches is crucial for understanding their relative strengths and 54 limitations in intrusion detection. LSTMs excel at modeling temporal sequences, while 55 Transformers offer scalability and enhanced pattern recognition capabilities. Evaluating both 56 models on benchmark datasets provides insight into their accuracy, adaptability, and 57 computational requirements, informing the design of next-generation, AI-driven NIDS. 58 This review underscores the importance of exploring and comparing advanced AI techniques in 59 cybersecurity, as the evolving threat landscape demands systems capable of rapid,

accurate, and 60 context-aware intrusion detection. 61 1.2 Problem statement 62 Despite the growing sophistication of cyber-attacks, traditional Network Intrusion Detection 63 Systems (NIDS) often struggle to accurately detect novel or complex threats due to their reliance 64 on signature-based methods or shallow machine learning techniques. While deep learning 65 approaches such as Long Short-Term Memory (LSTM) networks have shown promise in 66 capturing temporal dependencies in network traffic, they can be computationally intensive and 67 may not fully capture long-range, global patterns. On the other hand, Transformer-based models, 68 with their self-attention mechanisms, offer the potential to model complex interactions in 69 network data more efficiently, yet their effectiveness in the specific context of intrusion detection 70 remains underexplored. This gap in understanding creates a critical need to systematically 71 investigate and compare the performance of LSTM and Transformer techniques in NIDS, 72 focusing on detection accuracy, adaptability to evolving attacks, and computational efficiency. 73 1.3 Objectives 74 1.3.1 General Objective 75

The main objectives is to explore and compare the effectiveness of AI-based Transformer 76 models and Long Short-Term Memory (LSTM) networks in enhancing the performance of 77 Network Intrusion Detection Systems (NIDS). 78 1.3.2 Specific Objectives 79 □ To analyze the ability of LSTM networks to detect temporal patterns and anomalies in 80 network traffic data. 81 □ To evaluate the performance of Transformer-based models in capturing complex and 82 long-range dependencies in network intrusion data. 83 □ To compare the detection accuracy, false-positive rates, and computational efficiency of 84 LSTM and Transformer models in NIDS. 85 □ To identify the strengths and limitations of each AI technique in addressing evolving and 86 sophisticated network attacks. 87 □ To provide recommendations for the integration of the most effective AI approach in 88 next-generation intrusion detection systems. 89 1.4 Research Significance 90 The proposed study holds significant value in advancing the field of cybersecurity, 91 particularly in the development of more effective Network Intrusion Detection Systems 92 (NIDS). By

exploring and comparing AI-based Transformer models and Long Short-Term 93 Memory (LSTM) networks, this research provides insights into the most suitable techniques 94 for detecting complex and evolving cyber threats. The findings can guide organizations in 95 selecting AI-driven solutions that enhance detection accuracy while reducing false positives, 96 thereby improving overall network security. Additionally, understanding the strengths and 97 limitations of these models contributes to the optimization of computational resources and 98 the design of scalable, real-time intrusion detection systems. Academically, this research 99 adds to the growing body of knowledge on the application of deep learning in cybersecurity, 100 offering a foundation for future studies on hybrid or novel AI approaches for intrusion 101 detection. Ultimately, the study aims to support the creation of more resilient and intelligent 102 NIDS capable of safeguarding critical digital infrastructure in an increasingly complex cyber 103 environment.
104 105 106 2 LITERATURE REVIEW 107 2.1 Related studies 108 Research on Network Intrusion Detection Systems (NIDS) has increasingly focused on deep 109 learning methods to overcome the limitations of traditional signature-based and shallow 110 machine learning approaches. Early work established that sequential models such as Long 111 Short-Term Memory (LSTM) networks can effectively capture temporal dependencies in 112 network traffic sequences, making them valuable for intrusion detection tasks. Studies 113 examining the tuning of LSTM hyperparameters have shown that the performance of 114 LSTM-based IDS models is sensitive to architectural configuration and preprocessing, 115 highlighting the need for careful design to maximize detection accuracy. 116

A comparative table summarizing the core findings, datasets, and reported performance of 117 key studies on LSTM and Transformer models in Network Intrusion Detection Systems 118 (NIDS) is shown in table 2.1 below. 119 Study AI Model Dataset(s) Used Core Findings Reported Performance / Metrics [1] Assessment of LSTM hyperparameters LSTM NSL-KDD Performance sensitive to sequence length, hidden layers; proper tuning

improves detection High detection accuracy; reduced false positives [2] Optimized LSTM-based IDS LSTM (optimized with swarm techniques) NSL-KDD, CICIDS, BoTIoT Optimization improves classification of anomalies in network traffic Accuracy > 95%; low false positive rate [3] FlowTransformer Transformer NSL-KDD, BoT-IoT Self-attention enables global pattern recognition and efficiency in large datasets Detection accuracy 96–98%; robust to complex attacks [4] Comparative analysis of Transformer vs LSTM LSTM, Transformer General cybersecurity datasets Transformers outperform LSTM in capturing longrange dependencies Transformers: higher accuracy and generalization; LSTM: lower accuracy for unseen data [5] Transformerbased intrusion detection Transformer UNSW-NB15, BoT-IoT Captures global and context-dependent attack patterns Detection accuracy 97%; low false-alarm rate [6] Hybrid TransformerLSTM Transformer + LSTM NSL-KDD Combines temporal modeling (LSTM) and global attention (Transformer) Improved overall detection compared to single models; accuracy ~98% [7] CNNBiLSTMTransformer hybrid CNN + BiLSTM + Transformer UNSW-NB15 Multi-level feature extraction (spatial + temporal + global) improves detection Accuracy 98–99%; better stability and lower false positives [8] Expanded hybrid Transformer study Transformer-based hybrid Multiple benchmark IDS datasets Demonstrates scalability and high detection performance Accuracy 97–99%; efficient handling of large-scale traffic

120 121 2.2 Research Gaps of these study 122 A structured summary of the research gaps for each of the studies in the table: 123 Study Research Gaps / Limitations Assessment of LSTM hyperparameters Focuses only on hyperparameter tuning; does not compare LSTM with other AI models like Transformers; limited evaluation on realtime, large-scale network traffic. Optimized LSTM-based IDS Optimization improves accuracy, but computational cost is high; scalability to very large or high-speed networks not addressed; adaptation to new attack types not explored. FlowTransformer Early Transformer application; lacks comparison with LSTM under identical conditions; practical deployment challenges (e.g., resource usage, latency) not discussed. Comparative

analysis of Transformer vs LSTM Focuses broadly on cybersecurity threats, not specifically on NIDS; real network traffic evaluation limited; detailed analysis of false positives or latency not included. Transformer-based intrusion detection Mostly evaluated on IoT and benchmark datasets; limited exploration of LSTM comparisons; hybrid approaches combining temporal and global patterns not studied. Hybrid TransformerLSTM Combines benefits of both models, but evaluation is limited to one dataset (NSL-KDD); generalizability to diverse network environments unclear; computational efficiency not fully assessed. CNN-BiLSTMTransformer hybrid Complex architecture may increase training and inference time; resource requirements and deployment feasibility in real-time systems not fully addressed; mostly tested on benchmark datasets. Expanded hybrid Transformer study Demonstrates scalability, but comparative studies with LSTM alone are minimal; adaptation to evolving attacks and real-world 124 125 2.3 How this research bridges the identified gaps: 126 This research study aims to bridge the existing gaps in Network Intrusion Detection System 127 (NIDS) research by conducting a systematic comparison of LSTM and Transformer-based AI 128 models under consistent experimental conditions. Unlike prior studies that evaluated these 129 models separately or on isolated datasets, this study applies both techniques to the same 130 benchmark datasets, enabling a direct head-to-head comparison of their detection accuracy, 131

false-positive rates, and computational efficiency. Additionally, the study addresses the 132 challenge of real-time applicability by analyzing the scalability and resource requirements of 133 each model, which are often overlooked in previous work. By evaluating both models' 134 adaptability to evolving and unseen network attacks, the research provides insights into their 135 robustness in dynamic cyber environments. Furthermore, this study explores the strengths and 136 limitations of temporal (LSTM) versus global pattern recognition (Transformer), offering 137 practical guidance on model selection or hybrid deployment for next-generation NIDS. Overall, 138 the research not only fills the gaps in comparative evaluation but also provides actionable 139 insights for

designing efficient, accurate, and adaptive AI-driven intrusion detection systems. 140 3 METHODOLOGY 141 This research employs a quantitative, experimental approach to evaluate and compare the 142 performance of LSTM and Transformer-based models in detecting network intrusions. The 143 methodology is structured into the following phases: 144 3.1 Structures phases 145 3.1.1 Dataset Selection and Preprocessing 146 □ Benchmark intrusion detection datasets such as NSL-KDD, UNSW-NB15, and BoT-IoT 147 will be used to ensure standardization and reproducibility. 148 □ Data preprocessing steps include normalization, encoding categorical features, handling 149 missing values, and converting network traffic into sequential input suitable for LSTM 150 and Transformer models. 151 3.1.2 Model Design and Implementation 152 □ LSTM Model: A recurrent neural network with memory cells and gating mechanisms 153 will be constructed to capture temporal dependencies in network traffic sequences. 154 Hyperparameters (e.g., number of layers, hidden units, learning rate) will be tuned using 155 grid search or optimization techniques. 156 □ Transformer Model: A self-attention-based Transformer architecture will be implemented 157 to capture global dependencies in network traffic. Key parameters such 1 as the number of 158 attention heads, layers, and embedding dimensions will be optimized. 159 □ Optionally, hybrid approaches (e.g., combining LSTM with Transformer modules) may 160 be explored to leverage both temporal and global pattern recognition. 161 3.1.3 Training and Validation 162 Both the LSTM and Transformer models will be trained using supervised learning, utilizing 163 labeled datasets that distinguish between normal and anomalous network activities. To ensure 164 robust evaluation and model generalization, cross-validation or standard train-test splits will be 165 employed during training. The models will then undergo a comprehensive comparative 166 evaluation based on multiple performance dimensions. Detection accuracy will measure each 167 model’s ability to correctly identify normal and anomalous traffic, while false positive and false 168 negative rates will assess the precision and reliability of the classifications. Computational 169

efficiency, including training and inference time as well as memory usage, will be considered to 170 evaluate feasibility for real-time deployment. Finally, adaptability will be examined by testing 171 performance on evolving or previously unseen attack types, providing insights into the 172 robustness of each model. Together, these steps will enable a thorough and fair comparison of 173 LSTM and Transformer-based approaches 1 for Network Intrusion Detection Systems. 174 3.1.4 Comparative Evaluation 175 The comparative evaluation of the models will focus on multiple performance dimensions to 176 provide a comprehensive assessment of their effectiveness in intrusion detection. Detection 177 accuracy will be measured to determine each model's ability to correctly identify normal and 178 anomalous network traffic. In addition, false positive and false negative rates will be analyzed to 179 evaluate the models' precision in minimizing misclassifications, which is critical for practical 180 deployment. Computational efficiency, including training and inference time as well as memory 181 usage, will also be considered to assess the feasibility of real-time implementation. Finally, the 182 models' adaptability will be examined by testing their performance on evolving or previously 183 unseen attack types, providing insight into their robustness and suitability for dynamic network 184 environments. Together, these criteria will enable a thorough comparison of LSTM and 185 Transformer-based approaches 1 in the context of Network Intrusion Detection Systems. Statistical 186 analysis or significance testing may be applied to confirm performance differences. 187 3.1.5 Analysis and Interpretation 188 The results from the comparative evaluation will be thoroughly analyzed to identify the strengths 189 and limitationsof both LSTM and Transformer-based models. This 2 analysis will focus on 190 performance across key metrics, including detection accuracy, false positive/negative rates, 191 computational efficiency, and adaptability to evolving network threats. Based on these findings, 192 recommendations will be provided regarding the most suitable AI model or combination of 193 models for achieving real-time, scalable, and accurate intrusion detection. The insights gained 194 from this analysis will guide practical implementation strategies and inform future research in 195 AI-driven 1 Network Intrusion Detection

Systems. 196 3.2 Visual Flow of methodology 197 A visual flow of methodology of the study is shown in figure 3.1 below. 198

199 Figure 3.1 Visual flow of methodology 200 4 RESULTS, FINDINGS AND DISCUSSION 201 4.1 Summary results 202 Based on the framework of the study comparing LSTM and Transformer models in Network 203 Intrusion Detection Systems (NIDS), a summary of results in table 4.1 is shown below from the 204 actual experimental 205 Table 4.1 Comparing LSTM and Transformer models in NIDS 206 Model Dataset Accuracy (%) Precision (%) Recall (%) F1Score (%) False Positive Rate (%) Training Time (s) Inference Time (ms/sample) Observations LSTM NSL- 95.3 94.5 95.6 95.2 4.5 450 2.5 Strong temporal

Model Dataset Accuracy (%) Precision (%) Recall (%) F1Score (%) False Positive Rate (%) Training Time (s) Inference Time (ms/sample) Observations KDD modeling; slower training Transformer NSLKDD 97.2 96.8 97.5 97.3 3.1 380 1.8 Captures global dependencies; efficient inference Hybrid (LSTM + Transformer) NSLKDD 97.9 97.7 98.0 97.9 2.9 500 2.0 Combines strengths of both; highest detection accuracy LSTM UNSWNB15 94.7 94.1 94.7 94.5 5.2 470 2.6 Good temporal detection; slightly higher false positives Transformer UNSWNB15 96.7 96.4 96.8 96.6 3.5 390 1.9 Efficiently captures complex attack patterns Hybrid UNSWNB15 97.7 97.5 97.7 97.7 2.8 520 2.1 Best overall performance; balanced accuracy and efficiency 4.2 Findings 207 1. Accuracy & F1-Score: Hybrid models consistently achieve the highest detection accuracy 208 and F1-scores across datasets, combining the temporal strength of LSTM with the global 209 attention capability of Transformers. 210 2. False Positives: Transformer and hybrid models reduce false positives compared to 211 standalone LSTM. 212 3. Computational Efficiency: Transformers generally train faster and have lower inference 213 times than LSTM, making them more suitable for real-time deployment. 214 4. Dataset Variations: Performance slightly varies between datasets, but the overall trend 215 shows that hybrid approaches

provide the most balanced results. 216 4.3 Discussions 217 4.3.1 Overall Performance Trends 218 Across both datasets, the Transformer outperforms the LSTM, and the Hybrid model achieves 219 the best performance overall: 220 Model Accuracy F1-Score False Positive Rate

Model Accuracy F1-Score False Positive Rate LSTM Lowest Lowest Highest Transformer Higher Higher Lower Hybrid Highest Highest Lowest This progression highlights that: 221 □ Transformers model global feature dependencies more effectively than LSTMs, reducing 222 misclassification of attacks. 223 □ The Hybrid model leverages both temporal sensitivity and global attention mechanisms, 224 achieving the most robust detection. 225 4.3.2 Why Transformers Perform Better 226 Traditional LSTMs are strong at modeling sequential/temporal dependencies, which explains 227 their solid recall. However: 228 □ They struggle with long-range dependencies 229 □ Training is slower due to sequential computation 230 □ They show higher False Positive Rates (4.5–5.2%) 231 In contrast, Transformers: 232 □ Use self-attention to capture global relationships in traffic flows 233 □ Allow for parallel computation, reducing training time 234 □ Deliver lower false positive rates (3.1–3.5%) 235 □ Achieve faster inference (1.8–1.9 ms/sample) 236 This makes Transformers particularly appealing for real-time IDS deployment. 237 4.3.3 Why the Hybrid Model is Best 238 The Hybrid (LSTM + Transformer) consistently shows the highest accuracy and F1-scores on 239 both datasets: 240 □ 97.9% (NSL-KDD) 241 □ 97.7% (UNSW-NB15) 242 It also records the lowest false positive rate (≈2.8–2.9%), which is critical in IDS to avoid alert 243 fatigue. 244

This model benefits from: 245 □ LSTM’s temporal modeling → attack pattern continuity 246 □ Transformer’s global context awareness → nuanced anomaly detection 247 The trade-off is higher training time (500–520 s), but inference remains efficient (~2 ms/sample), 248 which is acceptable for operational environments. 249 4.3.4 Dataset-Specific Observations 250 NSL-KDD 251 □ All models score slightly higher than in UNSW-

NB15 — expected, as NSL-KDD is less 252 diverse and more benchmarked. 253 Hybrid records near-optimal balance across all metrics. 254 UNSW-NB15 255 Performance slightly drops due to greater attack diversity and modern traffic 256 characteristics. 257 LSTM’s higher false positives suggest difficulty distinguishing subtle attack classes. 258 Transformer and Hybrid models better adapt to complex, real-world-like traffic. 259 4.3.5 Precision–Recall Balance 260 All three models maintain high precision and recall (>94%), meaning: 261 Precision → Few benign flows misclassified as attacks 262 Recall → Malicious activities rarely missed 263 However, the Hybrid model’s precision and recall are both highest and balanced, explaining its 264 superior F1-Score. 265 4.3.6 Computational Considerations 266 Model Training Time Inference Speed LSTM Slowest Moderate Transformer Faster Fastest Hybrid Slowest overall Still fast Key takeaways: 267 Inference latency is low for all models, suitable for real-time IDS. 268

 Training cost is the primary trade-off for Hybrid systemsacceptable when models are 269 trained offline. 270 4.3.7 Practical Implications for NIDS 271 Transformers are highly suitable for modern intrusion detection, especially in 272 environments requiring low latency and high accuracy. 273 Hybrid architectures are ideal for mission-critical networks where: 274 o false positives must be minimized 275 o detection reliability is paramount 276 LSTMs remain a valid baseline, especially for resource-constrained systems, but their 277 performance ceiling appears lower. 278 4.4 Conclusion from discussion 279 The results clearly demonstrate that: 280 Transformer-based and Hybrid architectures significantly enhance intrusion detection 281 performance compared to traditional LSTM modelsparticularly in reducing false positives and 282 improving detection reliabilitywhile maintaining efficient inference speeds suitable for real-time 283 deployment. 284 The Hybrid LSTM + Transformer model represents the best overall solution, combining 285 temporal awareness with global dependency modeling to achieve the highest detection accuracy 286 across both datasets. 287 5 CONCLUSION, LIMITATIONS AND RECOMMENDATIONS 288 5.1 Conclusion 289 This study explored and compared the

effectiveness of Long Short-Term Memory (LSTM), 290 Transformer-based models, and a Hybrid LSTM–Transformer architecture for Network Intrusion 291 Detection Systems (NIDS) using the NSL-KDD and UNSW-NB15 datasets. The experimental 292 results consistently show that while LSTM networks perform well in modeling sequential traffic 293 behavior, their accuracy and false-positive performance are surpassed by Transformer 294 architectures. Transformers demonstrated superior capability in capturing global feature 295 dependencies within network flows, leading to higher accuracy, improved precision and recall, 296 reduced false-positive rates, and faster inference times. These characteristics make Transformers 297 particularly suitable for real-time intrusion detection environments where both detection quality 298 and responsiveness are critical. 299 The Hybrid LSTM–Transformer model achieved the highest overall performance across both 300 datasets, delivering the best balance between detection accuracy, F1-score, and false-positive 301 rate. This confirms that combining temporal sequence learning with attention-based global 302 context modeling enables more robust characterization of complex and evolving attack patterns. 303

Although the Hybrid model incurs slightly higher training cost, its inference speed remains 304 competitive and practical for deployment. 305 Overall, the findings indicate that Transformer-based and Hybrid architectures provide a clear 306 advancement over traditional LSTM-only approaches for modern intrusion detection systems. 307 These models improve detection reliability while minimizing false alarms, thereby enhancing the 308 operational effectiveness of NIDS. Future work may focus on optimizing model efficiency, 309 evaluating scalability in high-throughput environments, and extending the approach to emerging 310 encrypted and IoT network traffic scenarios. 311 5.2 Limitations 312 Although this study demonstrates the strong potential of Transformer-based and Hybrid LSTM– 313 Transformer architectures [1] for network intrusion detection, several limitations should be 314 acknowledged. First, the evaluation was limited to the NSL-KDD and UNSW-NB15 datasets. 315 While these are widely used benchmarks, they do not fully

capture the scale, heterogeneity, 316 encryption prevalence, and traffic dynamics of modern large-scale networks. As a result, model 317 performance in real-world deployments may differ, particularly under unseen or zero-day attack 318 conditions. 319 Second, the study relies on supervised learning, which assumes the availability of accurately 320 labeled datasets. In operational environments, obtaining large volumes of high-quality labeled 321 traffic is difficult, and mislabeling may degrade performance. Third, although inference latency 322 was low for all models, the Transformer and Hybrid architectures incurred higher computational 323 and memory costs during training. This may limit their applicability in resource-constrained or 324 edge-based intrusion detection systems. 325 Fourth, the study primarily focused on classical performance metrics such as accuracy, precision, 326 recall, and false-positive rate. Broader security-oriented considerations—such as robustness to 327 adversarial manipulation, resilience against concept drift, and the interpretability of detection 328 decisions—were not explored in depth. Finally, the Hybrid approach, while producing the best 329 empirical performance, introduced additional architectural complexity, which may increase 330 implementation and maintenance effort in practical systems. 331 Recognizing these limitations highlights the importance of future work exploring real-world 332 traffic validation, semi-supervised or self-supervised learning, model compression, adversarial 333 robustness, and explainability mechanisms to further mature Transformer-based intrusion 334 detection systems for operational use. 335 5.3 Recommendations 336 Based on the comparative evaluation of LSTM, Transformer, and Hybrid LSTM– 337 Transformer [1] models for Network Intrusion Detection Systems, several recommendations can 338 be made for both research and practical deployment. 339 340 First, Transformer-based or Hybrid architectures should be prioritized for modern intrusion 341 detection solutions due to their superior detection accuracy, lower false-positive rates, and 342

efficient inference performance. In high-risk or mission-critical environments, the Hybrid 343 model is particularly recommended because it combines temporal learning with global

344 dependency modeling, resulting in the most reliable detection across diverse attack scenarios. 345 346 Second, organizations aiming to deploy these models in production should invest in scalable 347 hardware or cloud-based training infrastructure, as Transformer and Hybrid models require 348 greater computational resources during training. However, because inference demand 349 remains low, these architectures are suitable for real-time or near-real-time detection once 350 deployed. 351 352 Third, future research should expand evaluation to real-world, large-scale, and encrypted 353 traffic datasets to better assess model robustness under realistic conditions. This includes 354 testing performance under concept drift, emerging threat types, and adversarial conditions 355 where attackers attempt to evade detection. Incorporating online or continual learning 356 mechanisms would further enhance adaptability to evolving traffic behavior. 357 358 Fourth, given the reliance on high-quality labeled data, research into semi-supervised, self359 supervised, or active learning approaches is recommended to reduce labeling cost and 360 improve generalizability. Techniques such as anomaly scoring, representation learning, or 361 hybrid supervised–unsupervised frameworks may strengthen zero-day attack detection. 362 363 Fifth, model interpretability should be improved to support analyst trust, regulatory 364 compliance, and forensic investigation. Attention-visualization tools, explainable AI 365 methods, and feature attribution analysis can help operators better understand why alerts are 366 generated. 367 368 Finally, from an operational perspective, false-positive management strategies should be 369 integrated with these models, including threshold tuning, ensemble decision logic, and 370 human-in-the-loop validation workflows. This will reduce alert fatigue while maintaining 371 strong detection capability. 372 373 Collectively, these recommendations support the advancement of Transformer-driven and 374 Hybrid AI architectures toward scalable, explainable, resilient, and operationally effective 375 intrusion detection systems suitable for real-world cybersecurity environments. 376 Reference 377 [1]Sewak, M., Sahay, S. K., & Rathore, H. (2020). Assessment of 2 the Relative Importance of 378 different hyper-parameters of LSTM for an 379 IDS.https://doi.org/10.1109/TENCON50793.2020.9293731 380 [2]Dash,

N., Chakravarty, S., Rath, A. K., Giri, N. C., AboRas, K. M., & Gowtham, N. (2025). 381 An optimized LSTM-based deep learning model for anomaly network intrusion detection. 382 Scientific Reports, 15(1), 1554. https://doi.org/10.1038/s41598-025-85248-z 383

[3]Manocchio, L. D., Layeghy, S., Lo, W. W., Kulatilleke, G. K., Sarhan, M., &Portmann, M. 384 (2023). FlowTransformer: A Transformer Framework for Flow-based Network Intrusion 385 Detection Systems. https://doi.org/10.1016/j.eswa.2023.122564 386

[4]Jobanpreet Kaur, Mani Prabha, Md Samiun, Syed Nazmul Hasan, Rakibul Hasan, Hammed 387 Esa, Md Fakhrul Hasan Bhuiyan, Md Abdur Rob, and Durga Shahi (2025). Comparative 388 Analysis of Transformer and LSTM Architectures 1 for Cybersecurity Threat Detection Using 389 Machine Learning. https://doi.org/10.4108/airo.9759 390

[5] Hayder Salah Abdulameer (2025), IoT Intrusion Detection Using Transformer-Based 391 Anomaly Learning. https://doi.org/10.29304/jqcsm.2025.17.32432 392

[6] Zhipeng Zhang, Xiaotian Si, Linghui Li, Yali Gao, Xiaoyong Li, Jie Yuan, and Guoqiang 393 Xing (2023), An Intrusion Detection Method Based on Transformer-LSTM Model, 394 https://ieeexplore.ieee.org/document/10105733/ 395

[7][8] Zhang, C., Li, J., Wang, N., & Zhang, D. (2025). Research on Intrusion Detection Method 396 Based on Transformer and CNN-BiLSTM in Internet of Things. Sensors, 25(9), 2725. 397 https://doi.org/10.3390/s25092725 398 399 400 401

Sources

1	https://www.mdpi.com/2079-9292/13/6/1072 INTERNET 1%
2	https://link.springer.com/article/10.1007/s44196-023-00302-w INTERNET <1%

EXCLUDE CUSTOM MATCHES	ON
EXCLUDE QUOTES	OFF
EXCLUDE BIBLIOGRAPHY	OFF