IJAICET
International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI : https://doi.org/10.68307/gzhs2192

Experimental Evaluation of a Trust-Aware Federated VAE–LSTM Framework for Intelligent Network Intrusion Detection Using Knowledge Quality Evaluation

Namrata Shrivastava[1], Pushparaj Shrivastava[2], Mahesh Kumar Bagwani[3]

[1]Sanjeev Agrawal Global Educational University, Bhopal, [2]LNCT University, Bhopal, [3]Sanjeev Agrawal Global Educational University, Bhopal

Abstract

The rapid expansion of cloud, Internet of Things (IoT), edge, and industrial networks has increased both the volume of network traffic and the diversity of cyber threats. Centralized intrusion detection can provide strong analytical capability but requires sensitive traffic to be transferred to a common repository, while conventional federated learning (FL) can still be affected by heterogeneous data, unreliable participants, and low-quality model updates. This paper presents an experimental evaluation of a Trust-Aware Federated Variational Autoencoder–Long Short-Term Memory (VAE–LSTM) framework that combines latent feature learning, sequential classification, client trust assessment, Knowledge Quality Evaluation (KQE), and FedProx aggregation. The VAE learns compact probabilistic representations of network-flow features and the LSTM models sequential attack behaviour. Before aggregation, client updates are evaluated using local accuracy, data quality, training consistency, and historical reliability. KQE further evaluates local accuracy, normalized loss, data diversity, and convergence stability. Experiments reported in the accompanying thesis use the CSE-CIC-IDS2018 dataset, five heterogeneous federated clients, 50 communication rounds, and comparisons with Decision Tree, Random Forest, SVM, CNN, LSTM, VAE–LSTM, FedAvg, and FedProx. The proposed framework reports 99.47% accuracy, 99.32% precision, 99.18% recall, 99.25% F1-score, and 0.998 ROC-AUC. It reaches 98% accuracy in 20 rounds and reduces total transmitted data from 925 MB to 370 MB under the reported setup. The results indicate that trust-aware and knowledge-quality-weighted aggregation can improve the reported detection performance, convergence behaviour, and communication efficiency while retaining the privacy-preserving property of federated learning.

Keywords—*Federated Learning, Network Intrusion Detection, VAE–LSTM, Knowledge Quality Evaluation, Trust-Aware Aggregation, FedProx, CSE-CIC-IDS2018.*

I. INTRODUCTION

A. BACKGROUND AND MOTIVATION

The rapid adoption of cloud computing, IoT, edge computing, and industrial cyber-physical systems has created highly distributed communication environments in which network traffic is generated by heterogeneous devices and services. Intrusion detection systems (IDSs) are therefore required to identify both known and previously unseen malicious behaviour while maintaining acceptable computational and communication costs. Recent surveys show that federated learning (FL) has become an important direction for intrusion detection because model training can be distributed without transferring raw traffic to a central repository [1], [2]. This property is particularly relevant when network traces are owned by independent organizations and cannot be freely pooled. However, privacy preservation alone does not guarantee trustworthy collaborative learning. Federated clients may hold non-IID traffic, differ in data quality, or produce unstable model updates. In a cybersecurity setting, an unreliable or compromised participant can therefore influence the global model in an undesirable direction. The broader FL literature identifies statistical heterogeneity, systems heterogeneity, communication cost, privacy leakage, and malicious participation as continuing challenges [2], [3]. FedProx was introduced to improve optimization under heterogeneous federated environments by adding a proximal constraint to local training [5]. These observations motivate an intrusion detection framework that evaluates not only the learned model but also the reliability and knowledge contribution of participating clients.

B. RESEARCH PROBLEM

The research problem addressed in this paper is the joint requirement for accurate intrusion detection, privacy-preserving collaborative training, and adaptive control of client contributions. Conventional machine-learning IDSs operate effectively when a representative centralized dataset is available, but centralized collection can expose sensitive network information. Standard FL reduces the need for raw-data exchange, yet equal or purely data-size-based aggregation may not adequately distinguish reliable clients from noisy or low-quality participants. Deep learning models can improve representation learning, but the resulting global model may still be affected by non-IID client distributions and unstable updates [1], [6].

The proposed research addresses this problem by combining four stages: probabilistic feature representation through a Variational Autoencoder (VAE), temporal classification through Long Short-Term Memory (LSTM), trust evaluation of client behaviour, and Knowledge Quality Evaluation (KQE) before FedProx aggregation. VAE provides a compact latent representation, while LSTM is used to capture temporal dependencies in network traffic [8], [9]. The resulting framework is evaluated experimentally rather than through a purely conceptual analysis.

C. OBJECTIVES AND CONTRIBUTIONS

The primary objective is to evaluate whether trust-aware and knowledge-quality-weighted aggregation improves the reported intrusion-detection performance and convergence behaviour of federated VAE–LSTM learning. The experimental objectives are to: (i) evaluate the proposed model against conventional machine-learning and deep-learning baselines; (ii) compare the proposed aggregation with FedAvg and FedProx; (iii) quantify trust and knowledge-quality scores across participating clients; (iv) measure communication and computational efficiency; and (v) examine statistical stability of the reported accuracy values.

IJAICET
International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI : https://doi.org/10.68307/gzhs2192

The main contributions are an integrated Trust Evaluation and KQE mechanism, a VAE–LSTM local learning architecture, adaptive aggregation using trust and knowledge quality with FedProx, and a consolidated experimental evaluation using the CSE-CIC-IDS2018 dataset. The contribution is therefore empirical as well as architectural: the paper reports how each component changes detection performance, convergence, communication volume, and client contribution within the experimental configuration described in the thesis.

II. RELATED WORK

A. FEDERATED LEARNING FOR INTRUSION DETECTION

Federated intrusion detection has developed from early demonstrations of privacy-preserving collaborative learning toward specialized architectures for IoT, industrial systems, and heterogeneous network environments. Agrawal et al. reviewed FL-based IDS research and identified privacy, non-IID data, communication overhead, and malicious participation as important challenges [1]. Lavaur et al. provided a structured survey and taxonomy of FL-based IDS architectures, emphasizing differences in domains, aggregation, and deployment models [2]. These studies establish FL as a useful architectural paradigm, but they also show that the quality of client updates remains an important research issue.

DeepFed demonstrated a federated deep-learning approach for industrial cyber-physical systems and combined deep neural learning with privacy-preserving collaborative training [6]. Other studies have investigated federated intrusion detection for wireless edge networks [16], clustered federated learning for industrial systems [12], and cyber-threat intelligence sharing [11]. More recent work has considered participant selection and heterogeneous data, indicating a transition from simple model aggregation toward mechanisms that explicitly manage client contribution [18].

B. DEEP REPRESENTATION LEARNING AND TEMPORAL MODELLING

Deep learning has become a standard approach for high-dimensional intrusion detection because neural networks can learn nonlinear feature relationships directly from network traffic [10]. Javaid et al. demonstrated an early deep-learning approach to network intrusion detection, while subsequent studies have explored CNN, recurrent, attention-based, and hybrid architectures [7]. LSTM is particularly relevant when attack behaviour is represented as a sequence because its gating mechanism allows information to be retained across multiple time steps [8]. VAE-based representation learning provides an alternative route for compressing correlated traffic features into a latent probabilistic space [9].

The VAE–LSTM combination used in this study follows this complementary division of responsibility. The VAE acts as the representation-learning stage, while the LSTM receives latent features for classification. The experimental comparison includes LSTM and VAE–LSTM separately so that the incremental effect of latent representation learning can be examined in addition to the effect of federated aggregation.

C. TRUST-AWARE AND KNOWLEDGE-AWARE AGGREGATION

Trust-aware FL attempts to distinguish reliable participants from clients whose updates may be noisy, unstable, or malicious. Recent work has explored trust mechanisms in intrusion detection and networked systems, including trust-based participant evaluation and robust aggregation [17], [19]. FedSBS similarly emphasizes participant selection for federated intrusion detection [18]. The proposed framework extends this direction by separating client reliability from knowledge quality: Trust Evaluation measures whether a client is reliable, whereas KQE measures how useful its learned knowledge is for the global model.

Recent federated intrusion-detection studies also report hybrid CNN–recurrent models, transformer-based approaches, and attention mechanisms [20]. These developments demonstrate the diversity of model architectures, but they do not by themselves resolve the aggregation problem. The experimental design in this paper therefore treats the aggregation mechanism as a first-class component and reports separate comparisons for FedAvg, FedProx, trust-based aggregation, and the combined Trust + KQE strategy.

III. PROPOSED TRUST-AWARE FEDERATED VAE-LSTM FRAMEWORK

A. OVERALL ARCHITECTURE

The proposed framework consists of five logical stages: local data preprocessing, VAE-based feature extraction, LSTM-based intrusion classification, client-level Trust and Knowledge Quality Evaluation, and trust-weighted FedProx aggregation. Each participating client retains its network-flow records locally. After preprocessing, the client trains the VAE–LSTM model and sends model parameters rather than raw network traffic to the aggregation server. The server evaluates the received updates before producing the next global model.

The architecture follows the experimental organization described in the thesis, where five virtual federated clients are connected to a central aggregation server. The non-IID client partition is intended to reproduce heterogeneous organizational traffic. The server first calculates trust values and then evaluates the knowledge quality of trusted updates. Final aggregation weights are derived from these scores before the FedProx update is distributed back to clients.

B. VAE FEATURE REPRESENTATION

The VAE transforms the preprocessed network-flow vector into a lower-dimensional latent representation. Its encoder estimates the parameters of a probability distribution, and the reparameterization operation allows latent samples to be generated while retaining differentiability. The decoder reconstructs the input during training, encouraging the latent space to preserve relevant traffic characteristics. This representation is subsequently supplied to the LSTM classifier.

The motivation for using the VAE is to reduce redundancy in high-dimensional flow features while retaining information useful for intrusion classification. In contrast to a deterministic autoencoder, the probabilistic latent representation can model variation in the input distribution. The thesis describes the VAE as the feature-extraction component and the LSTM as the

IJAICET
International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI : https://doi.org/10.68307/gzhs2192

classification component, forming a hybrid model that combines representation learning with sequential modelling.

C. TRUST EVALUATION AND KNOWLEDGE QUALITY EVALUATION

For client i, the thesis defines a trust score as a weighted combination of local accuracy, training consistency, data quality, and historical reliability. The coefficients sum to one, and higher values indicate stronger client reliability.

$$T_i = \alpha A_i + \beta C_i + \gamma D_i + \delta H_i$$

$$K_i = \lambda_1 A_i + \lambda_2 (1 - L_i) + \lambda_3 D_i + \lambda_4 C_i$$

This two-stage design distinguishes behavioural reliability from knowledge usefulness. A client can be reliable but contribute limited information if its local data are narrow, while a client with diverse data can still be unsuitable if its training behaviour is unstable. The final aggregation therefore gives greater influence to clients that score well on both dimensions. The reported client-level results show that the highest-scoring client also receives the largest final aggregation weight.

D. FEDERATED OPTIMIZATION

The aggregation stage uses FedProx to address heterogeneous local data and optimization behaviour. FedProx introduces a proximal term that constrains local model drift relative to the current global model [5]. In the proposed framework, this optimization is combined with adaptive aggregation weights derived from trust and KQE. The resulting process can be expressed conceptually as a weighted global update in which each client model is scaled by a reliability-and-quality factor before the FedProx aggregation step.

The experimental configuration contains five clients, 50 communication rounds, five local epochs, 100% client participation, and Adam as the local optimizer. The proposed framework is compared with FedAvg and FedProx under the same reported communication setting. This makes it possible to separate the effect of the proximal optimization from the additional trust and KQE mechanisms.

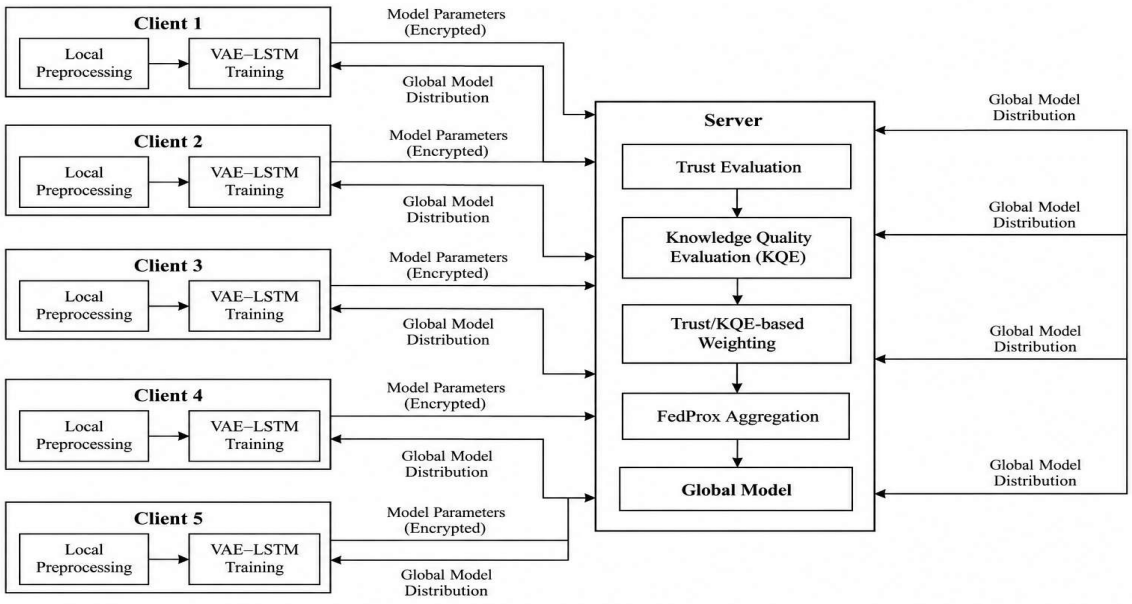


Fig. 3.1. Proposed Trust-Aware Federated VAE–LSTM Architecture

IV. EXPERIMENTAL SETUP

A. DATASET AND DATA PARTITIONING

The experiments use the CSE-CIC-IDS2018 dataset, developed by the Communications Security Establishment Canada and the Canadian Institute for Cybersecurity. The thesis describes the dataset as containing benign traffic and multiple attack categories, including brute force, DDoS, DoS, botnet, web attacks, infiltration, SQL injection, XSS, FTP/SSH brute force, and Heartbleed. Network flows are represented using statistical features generated by CICFlowMeter. The experimental data are treated as a multi-class intrusion-detection problem after preprocessing.

The reported split is 70% for training, 15% for validation, and 15% for testing. The training environment is then distributed across five virtual clients, with each client receiving a non-IID partition. This arrangement is important because the objective is not merely to classify a centralized test set, but to examine whether the proposed aggregation can learn from heterogeneous local traffic distributions.

B. IMPLEMENTATION ENVIRONMENT

The implementation uses Python with PyTorch for deep learning and Flower for federated learning. The reported software environment also includes Pandas, NumPy, Scikit-learn, Matplotlib, and Jupyter Notebook/Google Colab/VS Code. The hardware configuration specifies an eight-core processor, 32 GB RAM, an NVIDIA RTX 3060-class GPU, 1 TB SSD, five virtual federated clients, and a single aggregation server. The reported environment is therefore sufficient for local deep-learning training and repeated federated communication rounds.

IJAICET
International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI : https://doi.org/10.68307/gzhs2192

C. EVALUATION METRICS

The evaluation considers Accuracy, Precision, Recall, F1-score, ROC-AUC, global loss, communication rounds, transmitted data, communication time, execution time, memory use, trust scores, and knowledge-quality scores. Accuracy measures overall correct classification, while precision reflects the proportion of predicted attacks that are correct. Recall measures the proportion of actual attacks detected, and F1-score balances precision and recall. ROC-AUC is used to summarize discrimination capability across classification thresholds. Operational metrics are included because an intrusion detector deployed in a distributed environment must satisfy more than predictive accuracy. A model that reaches high accuracy but requires excessive communication or unstable convergence can be difficult to deploy. Accordingly, the reported experiments examine convergence, network overhead, computational cost, and client-level contribution alongside classification performance.

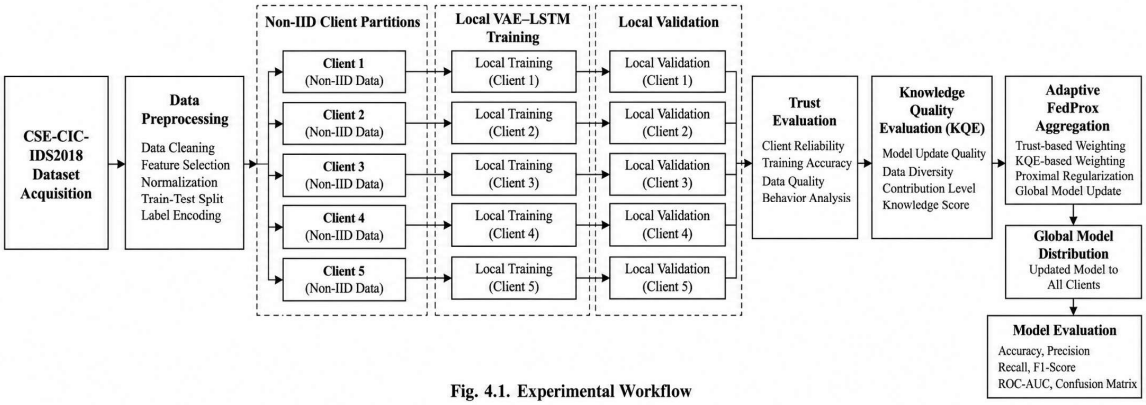


Fig. 4.1. Experimental Workflow

V. RESULTS AND ANALYSIS

A. OVERALL CLASSIFICATION PERFORMANCE

The main comparative experiment evaluates Decision Tree, Random Forest, SVM, CNN, LSTM, VAE–LSTM, FedAvg, FedProx, and the proposed framework. The reported results show a progressive improvement from conventional machine-learning models to deep-learning and federated approaches. The proposed framework records 99.47% accuracy, 99.32% precision, 99.18% recall, 99.25% F1-score, and 0.998 ROC-AUC. Among the reported federated baselines, FedProx reaches 98.62% accuracy and 98.30% F1-score, while FedAvg reaches 97.81% accuracy and 97.48% F1-score.

The results suggest that the gain is not attributable to a single architectural component. VAE–LSTM alone reaches 98.36% accuracy, showing the contribution of representation learning and temporal classification. The further increase under federated learning and then under trust + KQE aggregation indicates that the aggregation mechanism is associated with an additional improvement in the reported experimental configuration. These values should be interpreted as results of the thesis implementation rather than as universal performance guarantees across other datasets or deployments.

B. CONFUSION-MATRIX ANALYSIS

The reported binary confusion-matrix summary contains 29,432 correctly predicted normal records, 168 false positives, 245 false negatives, and 36,155 correctly detected attack records. These values indicate that both normal and malicious traffic are being classified with relatively small error counts in the reported test configuration. The false-negative count is particularly relevant to intrusion detection because undetected attacks can have direct security implications.

The confusion-matrix results should nevertheless be considered together with the multi-class evaluation because a binary aggregation of attack categories can conceal class-specific weaknesses. The thesis therefore complements the confusion matrix with precision, recall, F1-score, and ROC-AUC, providing a broader view of classification behaviour.

		Predicted Class: Benign (Normal Traffic)	Predicted Class: Attack (Intrusion Traffic)
Actual Class	Benign (Normal Traffic)	True Negative (TN): Benign traffic correctly classified as benign	False Positive (FP): Benign traffic incorrectly classified as attack
	Attack (Intrusion Traffic)	False Negative (FN): Attack traffic incorrectly classified as benign	True Positive (TP): Attack traffic correctly classified as attack

Fig. 5.1. Confusion-Matrix Interpretation

C. CONVERGENCE PERFORMANCE

Convergence is a major differentiator between the proposed method and conventional federated baselines. At communication round 20, the reported global accuracy is 98.12% for the proposed framework, compared with 94.03% for FedAvg and 95.47% for FedProx. At round 50, the corresponding values are 99.47%, 97.81%, and 98.62%. The proposed method therefore reaches the 98% accuracy level in 20 rounds, whereas the reported comparison requires 46 rounds for FedAvg and 31 rounds for FedProx.

The global loss values also show faster reduction. At round 20, the reported loss is 0.108 for the proposed method, compared with 0.231 for FedAvg and 0.198 for FedProx. By round 50, the values are 0.024, 0.073, and 0.058, respectively. These results are consistent with the intended role of trust and KQE:

IJAICET
International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI : https://doi.org/10.68307/gzhs2192

unreliable or less informative updates receive lower influence, while FedProx reduces local drift under heterogeneous data.

D. TRUST AND KNOWLEDGE QUALITY ANALYSIS

The client-level analysis shows trust scores of 0.981, 0.964, 0.951, 0.918, and 0.887 for Clients 1 through 5, respectively. The corresponding KQE scores are 0.985, 0.973, 0.954, 0.924, and 0.897. Final aggregation weights are 21.42%, 20.73%, 20.08%, 19.14%, and 18.63%. The reported values demonstrate that the framework does not treat the five clients as identical contributors.

The highest-scoring client has 99.14% local accuracy, a training loss of 0.021, data-diversity score of 0.98, and convergence-stability score of 0.99. The lowest-scoring client has 96.83% local accuracy, loss of 0.061, diversity of 0.89, and stability of 0.91. The observed relationship is consistent with the design of KQE, which rewards accurate, diverse, and stable local learning. The table does not establish that the client was malicious; it only shows that its measured contribution was lower under the defined scoring mechanism.

E. COMMUNICATION EFFICIENCY

The proposed framework reaches stable convergence in 20 communication rounds, whereas the reported FedAvg and FedProx comparisons use 50 rounds. With a model size of 18.5 MB per round, the total transmitted data are reported as 370 MB for the proposed framework versus 925 MB for each baseline, corresponding to a 60% reduction relative to the baseline totals. Average communication time per round is reported as 4.64 s for the proposed framework, compared with 4.85 s for FedAvg and 4.72 s for FedProx.

The reduction in total communication time is therefore primarily associated with fewer rounds rather than a large difference in per-round transmission time. The thesis reports total communication times of 92.80 s, 242.50 s, and 236.00 s for the proposed framework, FedAvg, and FedProx, respectively. Average network delay is reported as 141 ms for the proposed method, compared with 192 ms for FedAvg and 178 ms for FedProx.

F. COMPUTATIONAL PERFORMANCE AND STATISTICAL STABILITY

The proposed framework reports 541.28 s of training time, 18.46 s of testing time, 559.74 s total execution time, and 1812 MB memory usage. FedProx reports 598.35 s training time, 20.32 s testing time, 618.67 s total execution time, and 1898 MB memory use, while FedAvg reports 624.73 s, 21.14 s, 645.87 s, and 1916 MB, respectively. Although the proposed framework contains additional trust and KQE processing, the lower number of communication rounds reduces total execution time in the reported experiment.

The thesis also reports repeated-run statistical analysis. The proposed framework has a mean accuracy of 99.48% with a standard deviation of 0.12. A paired t-test reports p-values below 0.05 for the listed baseline comparisons, while one-way ANOVA reports $F = 318.92$ with $p < 0.001$. These statistics support the consistency of the reported experimental differences within the tested runs, but they do not by themselves establish generalization to other datasets, client populations, or attack conditions.

VI. DISCUSSION

A. INTERPRETATION OF THE EXPERIMENTAL FINDINGS

The results show a consistent relationship between architectural depth, federated optimization, and adaptive client weighting. Conventional machine-learning baselines remain computationally lightweight but show lower reported classification performance. CNN and LSTM improve the predictive metrics, while VAE–LSTM provides an additional improvement through latent representation learning. FedAvg and FedProx preserve the privacy-oriented distributed learning setup but differ in convergence and final accuracy. The proposed trust + KQE strategy records the highest reported values across the principal classification metrics.

The most important experimental observation is that client evaluation affects both model quality and communication behaviour. Trust scores provide a reliability-oriented filter, while KQE distinguishes the usefulness of the knowledge produced by trusted clients. This is consistent with the broader literature, where federated IDS research increasingly considers client selection, heterogeneity, and reliability rather than treating all updates as equally informative [1], [2], [18].

B. COMPARISON WITH EXISTING RESEARCH

The proposed framework is positioned relative to prior federated IDS research rather than treated as a universal replacement for other architectures. DeepFed demonstrated privacy-preserving federated deep learning for industrial cyber-physical systems [6], while later studies explored efficient federated IDSs, IIoT intrusion detection, and cyber-threat intelligence sharing [11], [14], [15]. Trust-aware research has additionally examined how client reliability can be incorporated into federated learning [17], [19]. The present framework combines these directions with a VAE–LSTM representation/classification pipeline and an explicit KQE stage.

The numerical results are dataset- and implementation-specific. The 99.47% accuracy reported here should therefore not be compared directly with numbers from studies using different datasets, label definitions, client counts, or evaluation protocols. The appropriate interpretation is that, under the experimental conditions reported in the thesis, the proposed method achieved the highest value among the included baselines.

C. PRIVACY, ROBUSTNESS, AND DEPLOYMENT CONSIDERATIONS

The federated architecture keeps raw network traffic at local clients and exchanges model parameters, which addresses the data-sharing requirement inherent in centralized training. However, model-parameter exchange does not automatically eliminate every privacy or security risk. Gradient or parameter leakage, poisoning, and backdoor attacks remain active research concerns in federated learning. The trust and KQE modules are intended to reduce the influence of unreliable updates, but the present thesis identifies adversarial-client evaluation as a limitation and does not claim comprehensive protection against all Byzantine or poisoning strategies.

For deployment, the reported communication reduction and lower total execution time are relevant to bandwidth-

IJAICET
International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI : https://doi.org/10.68307/gzhs2192

constrained distributed environments. Nevertheless, real deployments would require additional controls such as secure aggregation, authenticated clients, encrypted communication, robust anomaly checks on updates, and evaluation across multiple independent datasets. These considerations define a practical path from the reported experimental prototype toward operational network security systems.

VII. LIMITATIONS AND FUTURE WORK

A. LIMITATIONS

The reported experiments have several limitations that should be considered when interpreting the results. First, the evaluation uses the CSE-CIC-IDS2018 dataset as the primary benchmark, so external validity across other traffic collections is not established. Second, the federated simulation contains five clients, which is useful for controlled experiments but does not represent the full scale of large enterprise or national networks. Third, the thesis notes that Byzantine and model-poisoning attacks were not extensively evaluated. Finally, the reported communication and execution measurements depend on the stated hardware, software, model size, and client-server configuration.

B. FUTURE RESEARCH DIRECTIONS

Future work should evaluate the framework on multiple independent intrusion-detection datasets and larger client populations. Robustness experiments should introduce controlled poisoning, backdoor, Byzantine, and sybil-style behaviours so that the trust mechanism can be assessed under explicit adversarial conditions. The trust and KQE weights can also be learned adaptively rather than fixed manually, and the framework can be extended with secure aggregation or differential privacy where the associated utility cost is acceptable.

Another direction is deployment on real edge and IIoT nodes, where resource constraints, intermittent connectivity, and asynchronous participation may affect convergence. Explainability can also be integrated so that security analysts receive feature-level or flow-level evidence for an alert. These extensions would broaden the experimental scope without changing the central research objective of trustworthy collaborative intrusion detection.

VIII. CONCLUSION

This paper presented an experimental evaluation of a Trust-Aware Federated VAE–LSTM framework using Knowledge Quality Evaluation for intelligent network intrusion detection. The framework combines probabilistic feature representation, temporal classification, client trust assessment, knowledge-quality scoring, and FedProx aggregation. Under the reported CSE-CIC-IDS2018 experimental configuration, the proposed model achieved 99.47% accuracy, 99.32% precision, 99.18% recall, 99.25% F1-score, and 0.998 ROC-AUC. It reached the reported 98% accuracy threshold in 20 communication rounds and reduced total transmitted model data from 925 MB to 370 MB compared with the 50-round baselines.

The results indicate that the integration of Trust Evaluation and KQE can improve the reported quality of federated aggregation when combined with VAE–LSTM and FedProx. The contribution is therefore a unified experimental framework in which predictive accuracy, client reliability, knowledge usefulness, convergence, communication cost, and computational efficiency are evaluated together. The findings should be interpreted within the stated dataset and experimental limits, and future validation on independent datasets and explicit adversarial-client scenarios remains necessary.

REFERENCES

[1] S. Agrawal, S. Sarkar, O. Aouedi, G. Yenduri, K. Piamrat, M. Alazab, S. Bhattacharya, P. K. R. Maddikunta, and T. R. Gadekallu, “Federated Learning for intrusion detection system: Concepts, challenges and future directions,” Computer Communications, vol. 195, pp. 346–361, 2022.

[2] L. Lavaur, M.-O. Pahl, Y. Busnel, and F. Autrel, “The evolution of federated learning-based intrusion detection and mitigation: A survey,” IEEE Trans. Netw. Service Manag., vol. 19, no. 3, pp. 2309–2332, 2022.

[3] P. Kairouz et al., “Advances and open problems in federated learning,” Foundations and Trends in Machine Learning, vol. 14, nos. 1–2, pp. 1–210, 2021.

[4] Q. Yang, Y. Liu, T. Chen, and Y. Tong, “Federated machine learning: Concept and applications,” ACM Trans. Intell. Syst. Technol., vol. 10, no. 2, pp. 1–19, 2019.

[5] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, “Federated optimization in heterogeneous networks,” Proc. Mach. Learn. Syst., vol. 2, pp. 429–450, 2020.

[6] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, “DeepFed: Federated deep learning for intrusion detection in industrial cyber–physical systems,” IEEE Trans. Ind. Informat., vol. 17, no. 8, pp. 5615–5624, 2021.

[7] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, “A deep learning approach for network intrusion detection system,” in Proc. EAI SecureComm, 2016, pp. 21–26.

[8] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[9] D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” in Proc. Int. Conf. Learning Representations (ICLR), 2014.

[10] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[11] M. Sarhan, S. Layeghy, N. Moustafa, and M. Portmann, “Cyber threat intelligence sharing scheme based on federated learning for network intrusion detection,” Journal of Network and Systems Management, 2023.

[12] V. Jayagopal, M. Elangovan, S. S. Singaram, et al., “Intrusion detection system in industrial cyber-physical systems using clustered federated learning,” SN Computer Science, 2023.

[13] A. Belenguer, J. A. Pascual, and J. Navaridas, “GoWFed: A novel federated network intrusion detection system,” Journal of Network and Computer Applications, 2023.

[14] J. Li, X. Tong, J. Liu, and L. Cheng, “An efficient federated learning system for network intrusion detection,” IEEE Systems Journal, 2023.

[15] J. Zhang, C. Luo, M. Carpenter, and G. Min, “Federated learning for distributed IIoT intrusion detection using transfer approaches,” IEEE Trans. Ind. Informat., 2023.

[16] Z. Chen, N. Lv, P. Liu, et al., “Intrusion detection for wireless edge networks based on federated learning,” IEEE Access, 2020.

[17] D. Javeed, M. S. Saeed, M. Adil, et al., “A federated learning-based zero trust intrusion detection system for Internet of Things,” Ad Hoc Networks, 2024.

[18] H. N. C. Neto, J. Hribar, I. Dusparic, N. C. Fernandes, and D. M. Mattos, “FedSBS: Federated learning participant selection method for intrusion detection systems,” Computer Networks, 2024.

[19] S. Sankeawthong, X. Cao, S. Zhao, and Y. Zhang, “TrustFed-IDS: A trust-aware federated hybrid MLP-LSTM framework for robust intrusion detection in wireless sensor networks,” Computer Networks, 2026.

[20] P. Selvam, P. Karthikeyan, S. Manochitra, et al., “Federated learning-based hybrid convolutional recurrent neural network for multi-class intrusion detection in IoT networks,” Discover Internet of Things, 2025.

IJAICET
International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies
ISSN: 3139-8413 | Vol. 1 No. 2 (2026) | Original Research Article | www.ijaicet.com | DOI : https://doi.org/10.68307/gzhs2192

[21] M. K. Bagwani, A. Gangwar, K. Vishwakarma and V. K. Tiwari, “Real-time signature-based detection and prevention of DDOS attacks in cloud environments,” International Journal of Science and Research Archive, vol 12, no 2, pp 2929–2935, 2024, doi: 10.30574/ijsra.2024.12.2.1608. Source

[22] A. Khare, N. Choudhary and M. K. Bagwani, “Trustworthy and Adaptive Intrusion Detection for Imbalanced IoT Traffic,” International Journal of Artificial Intelligence, Cybersecurity and Emerging Technologies, vol 1, no 1, pp 80–86, 2026, doi: 10.68307/k9pxqe13. Source

[23] P. G. K. S. Mahesh Kumar Bagwani, "Performance comparison of REST API and GraphQL in a micro services architecture," International Conference on Data Science, Artificial Intelligence and Advanced Computing, 2024.