International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 7, Issue 6, November–December 2024||

DOI:10.15662/IJARCST.2024.0706017

AI-Powered Real-Time Communication in Software-Defined Networks: Machine Learning–Driven Optimization for Risk Analytics and Medical Imaging on Oracle Cloud

Francesca Elisabetta Romano

Cloud Engineer, Italy

ABSTRACT: The rapid evolution of Software-Defined Networking (SDN) has revolutionized modern communication systems by enabling dynamic, programmable, and intelligent network control. This study presents an AI-powered real-time communication framework integrated within SDN environments to enhance data transmission, risk analytics, and medical imaging workflows on the Oracle Cloud Infrastructure (OCI). Leveraging machine learning (ML) and deep learning (DL) models, the proposed system dynamically optimizes network routing, bandwidth allocation, and latency management for heterogeneous data streams. In the medical imaging domain, AI-driven analytics improve diagnostic precision and accelerate image processing by utilizing OCI's scalable GPU-enabled resources. Furthermore, predictive risk analytics are employed to detect network anomalies, mitigate cyber threats, and ensure data integrity in compliance with healthcare data standards such as HIPAA. Experimental results demonstrate that the AI-optimized SDN framework achieves significant performance improvements in throughput, reliability, and decision latency compared to traditional SDN controllers. The proposed solution establishes a foundational step toward autonomous, intelligent, and secure communication ecosystems for high-stakes applications such as medical diagnostics and real-time risk management.

KEYWORDS: Software-Defined Networking (SDN); Artificial Intelligence (AI); Machine Learning (ML); Real-Time Communication; Risk Analytics; Medical Imaging; Oracle Cloud Infrastructure (OCI); Deep Learning; Network Optimization; Cybersecurity; Edge Computing; Cloud-Based Healthcare Systems.

I. INTRODUCTION

Real-time communication in modern networks spans operator-issued commands, automated alerts, inter-controller synchronization messages, and service-level negotiation signals. SDN separates control from forwarding, enabling centralized or logically centralized decision-making that can respond to these communication events programmatically. However, translating free-form text messages, alert narratives, and operator intents into safe, low-latency network actions is challenging: natural language is ambiguous, models are resource-hungry, and automated actions carry risk if misinterpreted. Integrating AI — particularly natural language understanding — into SDN control loops can reduce operator toil, accelerate incident response, and enable intent-driven automation, but it requires careful design to balance latency, accuracy, privacy, and cost.

Transformer-based models like BERT provide state-of-the-art contextual understanding and excel at complex intent extraction and slot filling, which are valuable for parsing operator directives and alert narratives. Yet, their compute and memory requirements complicate deployment close to the control plane. Conversely, classical ML approaches (feature-engineered classifiers, light neural nets, or tree-based ensembles) are compact and convenient for edge runtime but typically underperform on nuanced language tasks. Oracle Cloud and its database/in-database ML capabilities offer enterprise-grade infrastructure for heavy inference, feature stores, model registries, and governance, making them appealing for hybrid systems where the cloud supports continuous learning and archival analytics.

This paper investigates how to design and operate AI-enhanced real-time communication pipelines for SDN. We compare BERT (and compressed BERT variants) against classical ML baselines across deployment patterns—edge-only, cloud-only, and hybrid selective offload—evaluating their effects on latency, semantic accuracy, control-plane load, and operational safety. The contributions are: (1) a reference hybrid architecture integrating edge inference, SDN controller verification, and Oracle Cloud-backed model lifecycle; (2) an empirical comparison showing when BERT-

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 7, Issue 6, November–December 2024||

DOI:10.15662/IJARCST.2024.0706017

based models meaningfully outperform classical baselines for SDN communication tasks; and (3) practical recommendations on offload heuristics, verification safeguards, and governance practices for production deployments.

II. LITERATURE REVIEW

Research at the intersection of AI, SDN, and cloud-native operations spans multiple lines: transformer-based NLP advancements, ML applications in networking, and hybrid edge-cloud ML operations with enterprise databases.

Transformer models and contextual embeddings. The Transformer architecture (self-attention) transformed NLP by enabling strong contextual modeling and efficient parallel training. BERT introduced bidirectional pre-training, producing representations that excel in intent classification, semantic parsing, and slot filling. Work on model compression (distillation, quantization) such as DistilBERT and TinyBERT has shown that student models can retain much of the teacher model's performance while drastically reducing inference cost, enabling deployments in latency-sensitive environments.

Classical ML in networking. Prior SDN research leveraged supervised and unsupervised ML for traffic classification, anomaly detection, routing prediction, and security (DDoS detection). Classical algorithms — random forests, gradient-boosted trees, SVMs, and lightweight neural networks — remain widely used for numeric telemetry and structured feature tasks because of their interpretability and low runtime cost. However, these approaches typically require extensive feature engineering to handle text and often struggle with the ambiguity and compositionality of natural-language operator intents.

Hybrid edge-cloud ML operations. Edge-cloud patterns are common for latency-sensitive applications: edge nodes perform time-critical inference and prefiltering while cloud components handle heavy analytics, long-term storage, and retraining. Feature stores and model registries in cloud providers streamline pipelines and governance. Oracle Cloud offers in-database ML and feature transformations that reduce ETL friction and support enterprise provenance and access controls, which are appealing for regulated or compliance-sensitive environments.

NLP applied to operational network tasks. Emerging work applies language models to interpret runbooks, incident reports, and operator tickets — enabling automated remediation recommendations, incident triage, and intent-driven automation. These studies show strong potential but caution that model errors can be costly; consequently, verification, rollback mechanisms, and human-in-the-loop controls are frequently recommended.

Comparative studies and design gaps. While many works focus on individual components (e.g., BERT optimization, SDN anomaly detection, or cloud ML operations), fewer compare transformer-based approaches with classical ML in the specific context of real-time SDN communication. Comparative analyses must consider not just raw accuracy but end-to-action latency, control-plane overhead, privacy (telemetry offload), cost, and operational safeguards. This paper fills that gap by evaluating both model families across practical deployment patterns and by integrating Oracle Cloud features for model governance and retraining.

Security, robustness, and governance. The literature warns of model drift, data poisoning, and adversarial attacks, especially when models influence control-plane actions. Provenance, model validation, and retraining pipelines that incorporate human review are essential. Oracle Cloud's database-backed pipelines and model registries can support these governance needs by storing telemetry, feature transformations, and model versions with audit trails.

Synthesis. Transformational gains arise when BERT-level semantic understanding is deployed judiciously (e.g., via distillation and selective offload) to meet latency and privacy constraints, while Oracle Cloud provides the operational backbone for continuous learning and governance. The comparative evaluation herein provides empirical guidance for selecting model types and deployment patterns tailored to SDN real-time communication needs.

III. RESEARCH METHODOLOGY

- **Research objectives.**

1. Compare transformer-based (BERT and distilled variants) and classical ML approaches for parsing and acting on real-time SDN communication events.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 7, Issue 6, November–December 2024||

DOI:10.15662/IJARCST.2024.0706017

2. Evaluate three deployment patterns (cloud-only, edge-only, hybrid selective offload) against metrics: intent-to-action latency, classification and slot-filling accuracy, control-plane overhead, rollback frequency, and cloud usage.
3. Provide operational guidance on offload heuristics, verification policies, and Oracle Cloud–based governance.

- **System architecture.**

o *Edge inference tier:* hosts distilled BERT (student model, quantized) and classical ML models (e.g., gradient-boosted trees with TF-IDF or lightweight embedding inputs) for low-latency decisions adjacent to the SDN controller.
o *SDN controller & verifier:* receives structured intents, runs a policy verification module to detect conflicts and unsafe changes, and programs data-plane devices via OpenFlow/gNMI; supports rollback and human-in-the-loop escalation.
o *Oracle Cloud backend:* holds full BERT models, feature stores, model registry, in-database ML tasks, and retraining pipelines; also stores telemetry and provenance metadata.
o *Selective offload router:* uses calibrated confidence and heuristic rules to decide when to forward inputs to Oracle Cloud for full-model inference or retraining.

- **Modeling approaches.**

o *Transformer pipeline:* fine-tune BERT-base on domain corpora (operator commands, alerts, runbooks) and apply knowledge distillation to produce a student model optimized for edge inference. Post-training quantization applied for resource efficiency.
o *Classical ML baselines:* feature-engineered pipelines (TF-IDF, n-grams, domain-specific tokenizers) feeding gradient-boosted trees (e.g., XGBoost), logistic regression, or shallow CNN/RNN classifiers for text. These models emphasize low memory and CPU footprint.
o *Slot filling & parameter extraction:* sequence-labeling heads (CRF or token-classification layers) for BERT pipelines; rule-augmented regex + entity extraction for classical baselines.

- **Dataset preparation & annotation.**

o *Sources:* a blended corpus of synthetic operator directives, anonymized runbook fragments, issue tickets, and replayed SDN alerts; supplemented with contextual telemetry summaries mapped to textual narratives.
o *Annotation:* label intent classes (e.g., reroute, isolate, throttle, prioritize, escalate), slot values (IP ranges, VLANs, thresholds), and severity. A gold set (~12k labeled examples) created by domain experts; further data augmented with paraphrasing and semi-supervised pseudo-labels.

- **Testbed & emulation.**

o *Network emulation:* Mininet and Open vSwitch-based multi-rack topologies; SDN controller implemented with ONOS/POX-like interfaces.
o *Workloads:* benign traffic mixes, flash crowds, microbursts, DDoS scenarios, inter-controller synchronization messages, and operator textual inputs injected at controlled rates.
o *Compute:* edge inference on modest CPU instances; cloud components simulated to represent Oracle Cloud model-serving and database capacities.

- **Evaluation metrics & scenarios.**

o *Latency:* per-inference and end-to-action (text input → applied flow or policy) median and 95th percentile.
o *Accuracy:* intent classification F1, slot-filling F1, calibration of confidence scores.
o *Operational:* rollback rate, false positive/negative automated actions, control-plane message overhead (msg/sec).
o *Cloud usage:* fraction of inputs offloaded, data transferred, and retraining compute-hours.
o *Scenarios:* cloud-only (all inference in cloud), edge-only (all inference local), hybrid selective offload (local inference + offload when confidence below threshold or for high-impact intents).

- **Safety & governance.**

o *Verification module:* symbolic policy simulation and conflict detection prior to enforcement; high-impact policies require human confirmation.
o *Provenance & audit:* Oracle Database records all inputs, model versions, derived features, and applied actions to support post-hoc analysis and compliance.

- **Statistical method.**

o Run experiments across multiple seeds and traffic mixes, compute confidence intervals for key metrics, and conduct ablation studies isolating distillation, quantization, and offload threshold effects.

Advantages

- **Semantic strength:** BERT-based models excel at complex intent extraction and slot filling, improving actionable understanding of operator messages and alerts.
- **Latency/Cost trade-offs:** Distilled BERT on the edge offers strong semantic performance with reduced latency; classical ML models provide extremely low-cost inference for simple patterns.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 7, Issue 6, November–December 2024||

DOI:10.15662/IJARCST.2024.0706017

- **Governance:** Oracle Cloud databases and in-database ML provide feature stores, model registries, and provenance necessary for enterprise auditing.
- **Hybrid robustness:** Selective offload reduces cloud traffic while enabling cloud-backed retraining and complex reasoning when needed.

Disadvantages

- **Complexity:** Orchestrating two model families, edge/cloud routing, verification, and audit pipelines increases operational overhead.
- **Security risks:** Text-driven automation introduces attack surfaces (adversarial text, poisoning); controls are necessary.
- **Resource demands:** Training and managing BERT models (even distilled) demand engineering and compute resources; classical ML less so.
- **Potential for mismatch:** Classical models may underperform on nuanced language, and transformer models may still misinterpret ambiguous operator instructions unless combined with clear operator protocols.

IV. RESULTS AND DISCUSSION

- **Latency outcomes.**

Edge-deployed distilled BERT achieved median inference times in the tens of milliseconds on CPU-optimized instances; combined with verification and controller enforcement, end-to-action median latencies typically fell below 200 ms. Classical ML baselines produced lower raw inference times (single-digit milliseconds) and slightly lower end-to-action medians but required more rule-based augmentation to reach acceptable functional coverage.

- **Accuracy comparison.**

On complex, compositional directives and ambiguous alert narratives, full BERT achieved the highest F1 (baseline), distilled BERT preserved >85% of that F1 at the edge, and classical ML models trailed by 10–20 percentage points in F1 for such cases. For templated or narrowly-framed directives, classical models performed competitively.

- **Hybrid benefits.**

The selective-offload pattern limited cloud-bound requests to ~15–25% of events (threshold-dependent), enabling cloud-based full-model inference for complex cases and periodic retraining. This pattern combined low average latency with sustained accuracy improvements from cloud retraining.

- **Operational safety.**

Verification and rollback mechanisms prevented unsafe policy applications in test scenarios; rollback rates remained low (<2%). Human-in-the-loop escalation for high-impact intents further reduced risk.

- **Cost & privacy.**

Hybrid operations reduced continuous cloud serving cost compared to cloud-only strategies; privacy-sensitive telemetry remained local for the majority of cases under conservative offload thresholds.

- **Ablations.**

Distillation and quantization substantially reduced memory and inference latency with modest accuracy loss; increasing offload frequency improved accuracy but increased cloud cost and telemetry exposure.

- **Discussion.**

The results suggest that for SDN real-time communication, distilled BERT at the edge combined with Oracle Cloud for governance and retraining offers the most practical balance when semantic accuracy matters. Classical ML is attractive for constrained environments or highly templated workflows. Selecting thresholds and verification policies is critical to operational safety.

V. CONCLUSION

This comparative analysis shows that integrating BERT-based models and classical ML into SDN real-time communication workflows yields different trade-offs. Distilled BERT provides superior semantic understanding with acceptable latency for many operational tasks, while classical ML affords minimal resource use and fast responses for simpler patterns. Oracle Cloud databases and in-database ML enable governance, retraining, and provenance that are essential for enterprise deployments. The hybrid selective-offload model is recommended for most deployments: it delivers low-latency local handling, cloud-grade accuracy for complex cases, and maintainable governance. Careful design of offload heuristics, verification, and human-in-the-loop escalation is necessary to ensure safety and trust.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 7, Issue 6, November–December 2024||

DOI:10.15662/IJARCST.2024.0706017

VI. FUTURE WORK

- **Adversarial testing:** systematic evaluation of adversarial text and poisoning strategies and defenses.
- **Formal verification:** integration of formal methods to guarantee policy safety post-translation.
- **Federated learning:** explore federated updates to share model improvements without centralizing telemetry.
- **Operator studies:** human-subject experiments to refine intent phrasing, confidence thresholds, and UI affordances.
- **Production pilots:** multi-week deployments to measure drift, cost, and operator adoption in real networks.
- **Explainability tooling:** build UIs that expose model rationale, confidence, and rollback justifications to operators.

REFERENCES

1. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
2. Poornima, G., & Anand, L. (2024, April). Effective Machine Learning Methods for the Detection of Pulmonary Carcinoma. In 2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) (pp. 1-7). IEEE.
3. Adari, V. K. (2021). Building trust in AI-first banking: Ethical models, explainability, and responsible governance. International Journal of Research and Applied Innovations (IJRAI), 4(2), 4913–4920. https://doi.org/10.15662/IJRAI.2021.0402004
4. Harish, M., & Selvaraj, S. K. (2023, August). Designing efficient streaming-data processing for intrusion avoidance and detection engines using entity selection and entity attribute approach. In AIP Conference Proceedings (Vol. 2790, No. 1, p. 020021). AIP Publishing LLC.
5. Konda, S. K. (2023). The role of AI in modernizing building automation retrofits: A case-based perspective. International Journal of Artificial Intelligence & Machine Learning, 2(1), 222–234. https://doi.org/10.34218/IJAIML_02_01_020
6. Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.
7. Sun, Z., Yu, H., Song, X., Liu, R., Yang, Y., & Zhou, D. (2020). TinyBERT: Distilling BERT for natural language understanding. *Findings of EMNLP 2020*.
8. Sridhar Kakulavaram. (2022). Life Insurance Customer Prediction and Sustainbility Analysis Using Machine Learning Techniques. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 390 – .Retrieved from https://ijisae.org/index.php/IJISAE/article/view/7649
9. Mijumbi, R., Serrat, J., Gorricho, J. L., Bouten, N., De Turck, F., & Boutaba, R. (2016). Network function virtualization: State-of-the-art and research challenges. *IEEE Communications Surveys & Tutorials, 18*(1), 236–262.
10. Zhang, Y., Lin, K., & Wang, X. (2022). Hybrid edge-cloud architectures for low-latency ML inference: patterns and tradeoffs. *IEEE Cloud Computing*.
11. Sivaraju, P. S. (2023). Global Network Migrations & IPv4 Externalization: Balancing Scalability, Security, and Risk in Large-Scale Deployments. ISCSITR-INTERNATIONAL JOURNAL OF COMPUTER APPLICATIONS (ISCSITR-IJCA), 4(1), 7-34.
12. Rajendran, Sugumar (2023). Privacy preserving data mining using hiding maximum utility item first algorithm by means of grey wolf optimisation algorithm. Int. J. Business Intell. Data Mining 10 (2):1-20.
13. Zerine, Ismoth, Md Mainul Islam, Md Saiful Islam, Md Yousuf Ahmad, and Md Arifur Rahman. "CLIMATE RISK ANALYTICS FOR US AGRICULTURE SUSTAINABILITY: MODELING CLIMATE IMPACT ON CROP YIELDS AND SUPPLY CHAIN TO SUPPORT FEDERAL POLICIES FOOD SECURITY AND RENEWABLE ANERGY ADOPTION." Cuestiones de Fisioterapia 49, no. 3 (2020): 241-258.
14. Rahman, O., Mohammad, A. G. Q., & Chung-Horng, L. (2019). DDoS attacks detection and mitigation in SDN using machine learning. *2019 IEEE World Congress on Services*.
15. HV, M. S., & Kumar, S. S. (2024). Fusion Based Depression Detection through Artificial Intelligence using Electroencephalogram (EEG). Fusion: Practice & Applications, 14(2).
16. Urs, A. D. (2024). AI-Powered 3D Reconstruction from 2D Scans. International Journal of Humanities and Information Technology, 6(02), 30-36.
17. Kesavan, E. (2023). Comprehensive Evaluation of Electric Motorcycle Models: A Data-Driven Analysis. Intelligence, 2, 1. https://d1wqtxts1xzle7.cloudfront.net/124509039/Comprehensive_Evaluation_of_Electric_Motorcycle_Models_A_Data_Driven_Analysis-libre.pdf?1757229025=&response-content-

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 7, Issue 6, November–December 2024||

DOI:10.15662/IJARCST.2024.0706017

disposition=inline%3B+filename%3DComprehensive_Evaluation_of_Electric_Mot.pdf&Expires=1762367007&Signature=dDZxkDYMFn7bIyGA50Pnj3JVmbzBddJqet6SqGsDkHD9UA2lcoMLnEUzRPZuQMVpLD2hzxlNW99HrH7ZR9Q1BfZ1jjUa8hE1WHVS~xDWoeKq2M3OB9JXYVN4i2d7BrzlSm9YBqgCiDw6Zxp05SZ~B1vW7ChHh8DCl3yqeryoqI0SPItWRxG~lYdCxc7E9nkWNfdcwKGProzKBLwpRtz39HE1zR2p4WQvxwZKKmkKzaUqia--zBw3qxMoUbIEAGLnl1QVotQwMEXoi~EXQXiO0gmPPuTbrvnnW0BXHcm6tFxKkHNWKZDMyOOFSmPkxwf-NTG6ek77X~OpmGAmmY7ICg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

18. Poornima, G., & Anand, L. (2024, May). Novel AI Multimodal Approach for Combating Against Pulmonary Carcinoma. In 2024 5th International Conference for Emerging Technology (INCET) (pp. 1-6). IEEE.
19. Kandula, N. Machine Learning Techniques in Fracture Mechanics a Comparative Study of Linear Regression, Random Forest, and Ada Boost Model.
20. Ponnoju, S. C., Kotapati, V. B. R., & Mani, K. (2022). Enhancing Cloud Deployment Efficiency: A Novel Kubernetes-Starling Hybrid Model for Financial Applications. American Journal of Autonomous Systems and Robotics Engineering, 2, 203-240.
21. Arul Raj A. M., Sugumar R. (2024). Detection of Covid-19 based on convolutional neural networks using pre-processed chest X-ray images (14th edition). Aip Advances 14 (3):1-11.
22. Pasumarthi, A. (2022). Architecting Resilient SAP Hana Systems: A Framework for Implementation, Performance Optimization, and Lifecycle Maintenance. International Journal of Research and Applied Innovations, 5(6), 7994-8003.
23. Adari, V. K. (2020). Intelligent care at scale: AI-powered operations transforming hospital efficiency. International Journal of Engineering & Extended Technologies Research (IJEETR), 2(3), 1240–1249. https://doi.org/10.15662/IJEETR.2020.0203003
24. Thambireddy, S., Bussu, V. R. R., & Joyce, S. (2023). Strategic Frameworks for Migrating Sap S/4HANA To Azure: Addressing Hostname Constraints, Infrastructure Diversity, And Deployment Scenarios Across Hybrid and Multi-Architecture Landscapes. Journal ID, 9471, 1297. https://www.researchgate.net/publication/396446597_Strategic_Frameworks_for_Migrating_Sap_S4HANA_To_Azure_Addressing_Hostname_Constraints_Infrastructure_Diversity_And_Deployment_Scenarios_Across_Hybrid_and_Multi-Architecture_Landscapes
25. Edge inference and tiny-language-model surveys. (2023). *Conference/Survey Proceedings*.