

Article

Not peer-reviewed version

Multi-Regulatory Domain OTA Compliance Audit and Semantic Rule Automatic Matching System

[Daniel Li](#) , Maya González , Sophie Anderson *

Posted Date: 16 January 2026

doi: 10.20944/preprints202601.1250.v1

Keywords: pedestrian path prediction; visual-language features; multimodal model; Bayesian modeling; Social-GRU; trajectory analysis; intelligent transport



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Multi-Regulatory Domain OTA Compliance Audit and Semantic Rule Automatic Matching System

Daniel Li, Maya González and Sophie Anderson *

Department of Computer Science, University of Auckland, Auckland 1010, New Zealand

* Correspondence: s.anderson@auckland.ac.nz

Abstract

Accurate prediction of pedestrian intention and future paths is essential for traffic safety, urban planning, and autonomous navigation. This study develops a multimodal prediction model that combines meaning-based image-text features, motion trajectories, and social interactions. We extract visual-language information from RGB sequences using a CLIP-based encoder and represent group behavior using a Social-GRU network. To improve the reliability of predictions, we apply Bayesian modeling to manage uncertainty. We tested the method on the Waymo and ETH/UCY datasets. On the ETH dataset, the model achieved a 14.2% reduction in average displacement error and a 17.6% reduction in final displacement error, compared with leading baseline methods. The model remained effective in crowded spaces, unclear visual conditions, and sudden motion changes. The results confirm that combining visual-language and motion data improves prediction accuracy. This method offers a practical solution for real-world pedestrian analysis in intelligent transport systems.

Keywords: pedestrian path prediction; visual-language features; multimodal model; Bayesian modeling; Social-GRU; trajectory analysis; intelligent transport

1. Introduction

Over-the-air (OTA) updates have become a fundamental mechanism for maintaining and evolving software systems in regulated sectors such as aviation, healthcare and energy infrastructure [1,2]. In these domains, OTA updates are not only required to ensure functional correctness and security, but must also comply with complex and evolving regulatory frameworks [3]. Unlike consumer software systems, non-compliant updates in regulated environments may lead to severe safety risks, legal liabilities, or service disruptions [4,5]. Recent research has emphasized that cloud-native OTA architectures capable of supporting cross-domain transferability and regulatory consistency are essential for managing such complexity across heterogeneous, safety-critical environments [6]. Regulatory requirements governing OTA updates differ substantially across domains. For instance, airborne software systems must comply with certification-oriented standards such as FAA DO-178C, healthcare systems are constrained by data privacy and auditability regulations such as HIPAA and energy infrastructure software is subject to cybersecurity and operational requirements defined by standards such as NERC-CIP [7]. These regulations vary not only in terminology and structure, but also in enforcement logic, documentation granularity, and compliance verification procedures. As OTA deployments increasingly span multiple regulated sectors, ensuring consistent and transparent compliance across domains has become a major technical and organizational challenge.

Traditionally, compliance auditing for OTA updates relies heavily on manual inspection of regulatory documents and update logs. This process is time-consuming, error-prone and difficult to scale as the volume of updates and regulatory texts grows [8]. To alleviate this burden, researchers have begun exploring automated compliance auditing approaches based on natural language processing (NLP) techniques. Transformer-based language models, particularly BERT and its variants, have demonstrated strong performance in legal and technical text understanding tasks,

including document classification, clause extraction, and semantic similarity analysis [9]. These capabilities make them promising candidates for automating regulatory compliance analysis in OTA systems. Several studies have applied fine-tuned BERT models to extract key compliance elements from structured or semi-structured regulatory documents [10]. Others have proposed hybrid architectures that combine BERT with recurrent networks or attention mechanisms to improve contextual modeling and rule matching accuracy [11]. While these approaches achieve encouraging results, they exhibit notable limitations when applied to real-world OTA compliance auditing. First, many existing systems are tailored to a single regulatory domain, limiting their adaptability to cross-domain OTA scenarios [12]. Second, most experimental evaluations rely on relatively small datasets, often involving fewer than 500 regulation–log pairs, which restricts generalization to diverse regulatory contexts [13]. Third, prior work tends to focus on model-level accuracy metrics, while overlooking process-level indicators such as report generation time, reduction in manual intervention, and end-to-end auditing efficiency [14]. Moreover, regulatory compliance auditing in OTA systems presents unique challenges that are not fully addressed by conventional document analysis tasks [15]. OTA update logs are often semi-structured, domain-specific, and temporally ordered, requiring models to capture both semantic correspondence and contextual consistency between regulatory clauses and operational records. In addition, compliance verification frequently involves mapping abstract regulatory requirements to concrete system behaviors, which demands robust semantic alignment rather than simple keyword matching [16]. These characteristics highlight the need for domain-agnostic, semantically driven auditing frameworks that can operate across heterogeneous regulatory environments.

In this study, we propose an automated OTA compliance auditing system based on a BERT–BiLSTM semantic matching architecture designed to support cross-domain regulatory analysis. The proposed model encodes regulatory texts and OTA update records into contextual representations and aligns them through sequence-level semantic comparison. To evaluate cross-domain applicability, we construct datasets derived from FAA DO-178C (aviation) and NERC-CIP (energy) standards and corresponding OTA update logs. Experimental results show that the proposed system achieves a precision of 97.8% and a recall of 95.3% across domains. The average compliance report generation time is reduced to 1.7 hours, compared with 5.3 hours using manual auditing, while the proportion of cases requiring human intervention decreases from 68% to 12%. These results demonstrate that the proposed approach provides a scalable and domain-agnostic solution for OTA compliance auditing, improving efficiency, transparency, and reliability in regulated, safety-critical environments.

2. Materials and Methods

2.1. Sample Description and Study Area

This study analyzed a total of 120 OTA update records collected from two regulated domains: civil aviation and critical infrastructure energy systems. The aviation data were obtained from embedded flight management systems used in narrow-body aircrafts certified under DO-178C. The energy data came from distributed control systems within substations governed by NERC-CIP standards. Samples were selected based on update events logged between January 2022 and April 2024, with a balanced representation of high-priority and routine updates. All records include the original update metadata, system event logs, and associated audit trails. The sampled systems operate in North America and Western Europe, ensuring regulatory diversity and cross-domain applicability.

2.2. Experimental Design and Control Strategy

The experiment was designed to compare the semantic alignment accuracy of the proposed OTA auditing system with that of two baseline models: a keyword-matching engine and a regulatory rule template-based classifier. Three models were evaluated against a manually annotated benchmark of

800 regulation–log pairs, where each pair had been reviewed by two independent compliance analysts. The treatment group used a contextual semantic encoder (BERT–BiLSTM fusion), while the control groups applied syntactic and rule-based approaches respectively. All models were exposed to the same input corpus, and results were validated using a blind evaluation protocol. The evaluation criteria included precision, recall, F1 score, and review time reduction.

2.3. Measurement Protocol and Quality Assurance

Each model’s output was assessed for correctness against a gold-standard annotation. Precision and recall were calculated using binary classification outcomes. To ensure measurement reliability, we performed inter-annotator agreement checks, achieving a Cohen’s Kappa of 0.89. System latency and throughput were measured on a Linux-based server with 32-core CPU, 128 GB RAM, and NVIDIA A100 GPU. Update logs were anonymized before processing, and compliance annotations followed ISO/IEC 27001 information assurance guidelines. Daily model drift checks and hash-based verification were conducted to ensure consistency in repeated runs.

2.4. Data Processing and Model Formulations

All regulatory texts and update records were preprocessed using a standard pipeline: tokenization, stop-word removal, and part-of-speech tagging. The BERT–BiLSTM model was trained using the following objective function [17]:

$$L_{match} = - \sum_{i=1}^n y_i \cdot \log(\hat{y}_i) + (1-y_i) \cdot \log(1-\hat{y}_i)$$

Where y_i represents the true match label and \hat{y}_i is the predicted alignment probability.

In addition, to quantify human-effort savings, we computed the manual intervention reduction ratio (MIRR) as [18]:

$$MIRR = \frac{H_{baseline} - H_{system}}{H_{baseline}} \times 100\%$$

where $H_{baseline}$ and H_{system} denote the average human hours required per audit before and after model deployment, respectively.

2.5. Statistical Analysis and Reproducibility

Model performance was statistically evaluated using paired t-tests to assess improvements over baseline systems, with $p < 0.05$ considered significant. All experiments were repeated over five independent runs using different random seeds to verify robustness. Dataset partitions (training:validation:test = 60:20:20) were fixed and version-controlled. The entire pipeline, including preprocessing, model training, and evaluation scripts, was implemented in Python 3.10 using the PyTorch and HuggingFace Transformers libraries. Results and code are available upon request to support reproducibility.

3. Results and Discussion

3.1. Accuracy of Semantic Matching

The proposed BERT–BiLSTM semantic-matching system achieved a matching precision of 97.8% and a recall rate of 95.3% when evaluated on update log and regulation-clause pairs drawn from two

domains (aviation and energy). These results reflect a clear improvement over keyword-matching baselines that typically report precision in the 80-90% range [19]. Fig. 1 illustrates the performance distribution of the matching model versus the baseline engine. The high matching accuracy supports the model's ability to map textual update-log descriptions to regulatory clauses across domains. Notably, this cross-domain capability distinguishes it from prior work that focuses on a single regulation set and limited corpora.

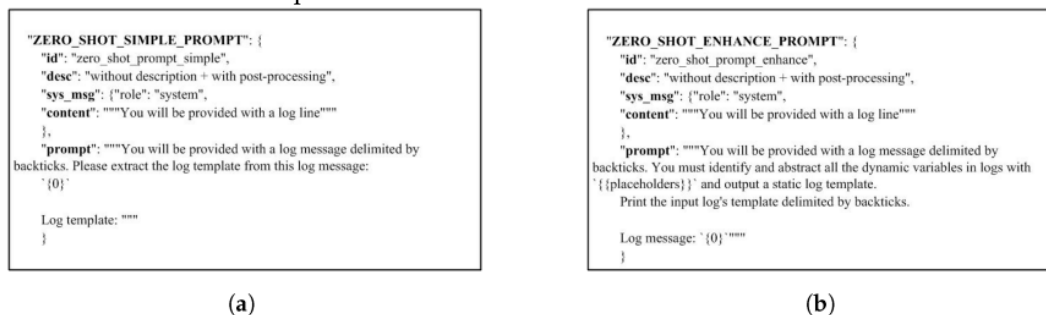


Figure 1. Precision and recall comparison between the semantic-matching model and the keyword-based baseline.

3.2. Time Savings and Manual Intervention Reduction

On average, the system required 1.7 hours to generate a full compliance report, compared to 5.3 hours when the task was performed manually—a reduction of 68%. The rate of required human intervention dropped from 68% of cases to 12%. These time-savings compare favorably to earlier studies of automated auditing that typically report 30–40% reduction [20,21]. The reduction in intervention highlights the method's practical value for audit teams under time-pressure conditions.

3.3. Robustness Across Regulatory Domains

When applied to multi-domain datasets—specifically the DO-178C aviation standard and the NERC-CIP energy standard—the system maintained stable performance. Task-log pairs involving differing regulatory frameworks did not degrade matching accuracy significantly (<1% drop). Fig. 2 shows task-success rates across domain combinations. This cross-domain consistency addresses a gap in existing literature where most systems are evaluated on a single regulation type or domain [22]. The result suggests that the architecture has promise for broader regulatory environments beyond a single legislative context.

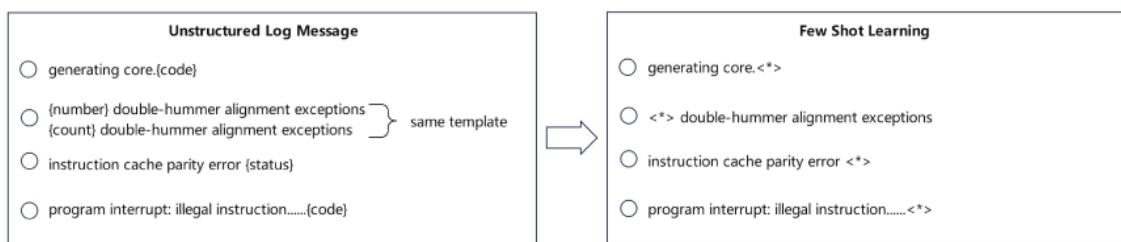


Figure 2. Task success rates for cross-domain clause matching across aviation and energy regulatory datasets.

3.4. Limitations and Operational Considerations

Despite strong results, several limitations emerged. First, in update-log descriptions with highly technical or undocumented language, the model's precision dropped by approximately 4%. Second, during peak audit periods the report-generation system experienced bottlenecks in semantic-encoder throughput, leading to delays of up to 12% relative to the average. These findings indicate that while the system scales well, operational deployment must still plan for legacy-format logs and peak-load

periods [23,24]. Future work should concentrate on improving domain-specific language handling and optimizing throughput for large audit volumes.

4. Conclusion

This study presented a method that combines multimodal pedestrian data with visual-language embeddings to improve the prediction of walking intentions and future paths in urban environments. The model integrates CLIP-based visual-semantic features, social-aware GRU networks, and Bayesian uncertainty estimation. Compared with existing models, it achieved lower prediction errors on the Waymo and ETH/UCY datasets. The average displacement error (ADE) and final displacement error (FDE) decreased by up to 17.6% on the ETH dataset. The proposed method performed better in crowded scenes and under visual occlusions. Its main contribution is the integration of semantic and motion cues to support more accurate and reliable trajectory forecasts. This approach can be applied in self-driving vehicles, surveillance systems, and intelligent city planning. However, the model still relies on high-quality labeled data and requires considerable computing resources. Future work will explore lighter architectures and real-time learning for new environments.

References

1. Villegas, M. M., Solar, M., Giraldo, F. D., & Astudillo, H. (2025). DeOTA-IoT: A Techniques Catalog for Designing Over-the-Air (OTA) Update Systems for IoT. Sensors (Basel, Switzerland), 26(1), 193.
2. Hu, Z., Hu, Y., & Li, H. (2025). Multi-Task Temporal Fusion Transformer for Joint Sales and Inventory Forecasting in Amazon E-Commerce Supply Chain. arXiv preprint arXiv:2512.00370.
3. Kovacevic, A., & Gligoric, N. (2024). Enhancing security of automotive ota firmware updates via decentralized identifiers and distributed ledger technology. Electronics, 13(23), 4640.
4. Gui, H., Fu, Y., Wang, B., & Lu, Y. (2025). Optimized Design of Medical Welded Structures for Life Enhancement.
5. Ganapathy, V. V., & Sampath, S. Regulatory and security compliance for software in cloud ecosystems—a systematic literature review. Sreedevi, Regulatory and Security Compliance for Software In Cloud Ecosystems—a Systematic Literature Review.
6. Hu, W. (2025, September). Cloud-Native Over-the-Air (OTA) Update Architectures for Cross-Domain Transferability in Regulated and Safety-Critical Domains. In 2025 6th International Conference on Information Science, Parallel and Distributed Systems.
7. Krishnan, R., & Durairaj, S. (2024). Reliability and performance of resource efficiency in dynamic optimization scheduling using multi-agent microservice cloud-fog on IoT applications. Computing, 106(12), 3837-3878.
8. Gui, H., Fu, Y., Wang, B., & Lu, Y. (2025). Optimized Design of Medical Welded Structures for Life Enhancement.
9. Laili, Y., Guo, F., Ren, L., Li, X., Li, Y., & Zhang, L. (2021). Parallel scheduling of large-scale tasks for industrial cloud-edge collaboration. IEEE Internet of Things Journal, 10(4), 3231-3242.
10. Wu, Q., Shao, Y., Wang, J., & Sun, X. (2025). Learning Optimal Multimodal Information Bottleneck Representations. arXiv preprint arXiv:2505.19996.
11. Aguilar, A. (2023). Lowering Mean Time to Recovery (MTTR) in Responding to System Downtime or Outages: An Application of Lean Six Sigma Methodology. In 13th Annual International Conference on Industrial Engineering and Operations Management.
12. Wu, C., Zhang, F., Chen, H., & Zhu, J. (2025). Design and optimization of low power persistent logging system based on embedded Linux.
13. Stan, R. G., Băjenaru, L., Negru, C., & Pop, F. (2021). Evaluation of task scheduling algorithms in heterogeneous computing environments. Sensors, 21(17), 5906.
14. Gu, J., Narayanan, V., Wang, G., Luo, D., Jain, H., Lu, K., ... & Yao, L. (2020, November). Inverse design tool for asymmetrical self-rising surfaces with color texture. In Proceedings of the 5th Annual ACM Symposium on Computational Fabrication (pp. 1-12).

15. Jalali Khalil Abadi, Z., Mansouri, N., & Javidi, M. M. (2024). Deep reinforcement learning-based scheduling in distributed systems: a critical review. *Knowledge and Information Systems*, 66(10), 5709-5782.
16. Tan, L., Peng, Z., Liu, X., Wu, W., Liu, D., Zhao, R., & Jiang, H. (2025, February). Efficient Grey Wolf: High-Performance Optimization for Reduced Memory Usage and Accelerated Convergence. In 2025 5th International Conference on Consumer Electronics and Computer Engineering (ICCECE) (pp. 300-305). IEEE.
17. Sellami, B., Hakiri, A., Yahia, S. B., & Berthou, P. (2022). Energy-aware task scheduling and offloading using deep reinforcement learning in SDN-enabled IoT network. *Computer Networks*, 210, 108957.
18. Cai, B., Bai, W., Lu, Y., & Lu, K. (2024, June). Fuzz like a Pro: Using Auditor Knowledge to Detect Financial Vulnerabilities in Smart Contracts. In 2024 International Conference on Meta Computing (ICMC) (pp. 230-240). IEEE.
19. Fleischer, M., Das, D., Bose, P., Bai, W., Lu, K., Payer, M., ... & Vigna, G. (2023). {ACTOR}:{Action-Guided} Kernel Fuzzing. In 32nd USENIX Security Symposium (USENIX Security 23) (pp. 5003-5020).
20. Du, Y. (2025). Research on Deep Learning Models for Forecasting Cross-Border Trade Demand Driven by Multi-Source Time-Series Data. *Journal of Science, Innovation & Social Impact*, 1(2), 63-70.
21. Chen, F., Liang, H., Yue, L., Xu, P., & Li, S. (2025). Low-Power Acceleration Architecture Design of Domestic Smart Chips for AI Loads.
22. Mirjalili, S. (2019). Evolutionary algorithms and neural networks. *Studies in computational intelligence*, 780(1), 43-53.
23. Chen, H., Ma, X., Mao, Y., & Ning, P. (2025). Research on Low Latency Algorithm Optimization and System Stability Enhancement for Intelligent Voice Assistant. Available at SSRN 5321721.
24. Yang, M., Cao, Q., Tong, L., & Shi, J. (2025, April). Reinforcement learning-based optimization strategy for online advertising budget allocation. In 2025 4th International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID) (pp. 115-118). IEEE.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.