Journal of Data Analytics and Intelligent Decision-making

Journal homepage: jdaid.qom.ac.ir

University of Qom

Video Coding Machine Architecture for Smart Urban Traffic Optimization with Deep Learning

Mehran Riki*[a], Fatemeh Mohammadi[b], Ehsan Eslami[C]

A. Department of Computer Engineering, Faculty of Electrical and Computer Engineering,Technical and Vocational University, Tehran, Iran.
B. Master's Student in Computer Science, Faculty of Mathematics, Statistics, and Computer Science, University of Sistan and Baluchestan, Zahedan, Iran.
C. Department of Computer Engineering, Faculty of Electrical and Computer Engineering,Velayat University, Iranshahr, Iran.

A R T I C L E I N F O

Keywords:
Intelligent Transportation Systems
Video Coding Machine, Versatile Video Coding
Traffic Congestion Prediction
CNN-RNN Hybrid Model
Smart Urban Traffic
Real-time Video Processing

A B S T R A C T

Intelligent Transportation Systems (ITS) are essential for modern urban infrastructure but grapple with real-time processing of voluminous traffic video data amid bandwidth and latency limitations. This paper introduces a novel Video Coding Machine (VCM) architecture that synergistically combines Versatile Video Coding (VVC) with an adaptive bitrate optimization algorithm—driven by neural features—and a hybrid Convolutional Neural Network (CNN)–Recurrent Neural Network (RNN) model for optimized compression and congestion prediction. The VVC core, enhanced by dynamic quantization parameter (QP) adjustments, minimizes data volume while upholding perceptual quality, whereas the CNN extracts spatial features (e.g., vehicle density) and the RNN captures temporal dynamics for precise forecasting. Evaluated on diverse real-world datasets (Cityscapes, BDD100K, Tehran traffic), the system attains 94% prediction accuracy (with 93% precision and 95% recall), 60% data reduction, and 25% faster processing versus baselines like H.264/AVC and H.265/HEVC. This framework delivers a scalable, efficient solution for smart cities, fostering real-time ITS applications, substantial cost efficiencies in storage/transmission, and improved urban mobility/safety. By bridging advanced compression and deep learning, it advances sustainable traffic management paradigms.

* Corresponding author.
E-mail addresses: mriki@tvu.ac.ir (M. Riki), FatemeMohammadi@pgs.usb.ac.ir (F. Mohammadi) , e.eslami@velayat.ac.ir (E. Eslami)

Received 30 June 2025; Received in revised form 6 July 2025; Accepted 12 August 2025, Available online 16 September 2025

Cite this article: Riki, M., Mohammadi, F., Eslami, E. (2025). Video Coding Machine Architecture for Smart Urban Traffic Optimization with Deep Learning .*Journal of Data Analytics and Intelligent Decision-making*, 1(3), 1-12.
https://doi.org/10.22091/jdaid.2025.14036.1003

1. Introduction

With the accelerating pace of urbanization and the surging number of vehicles in metropolitan areas, Intelligent Transportation Systems (ITS) have emerged as a critical solution for addressing traffic challenges and enhancing urban mobility. The primary objectives of ITS include dynamic traffic flow management, improved travel safety, reduced commute times, optimized fuel efficiency, and an overall elevated commuting experience for citizens. Video data, captured by extensive urban camera networks, serves as a rich informational resource in ITS, providing visual insights into traffic conditions, vehicle behaviors, incidents, and patterns. Real-time analysis of these videos enables predictive congestion forecasting, bottleneck identification, adaptive traffic signal control, violation detection, and route recommendations to drivers. However, the large volume of video data poses substantial hurdles, including limitations in bandwidth within communication networks, exorbitant storage costs, intensive computational demands for real-time processing, and stringent latency requirements in ITS applications.

Recent advancements in deep learning have significantly propelled ITS capabilities, particularly in traffic prediction and analysis. Transformers have gained prominence for their superior handling of sequential dependencies in traffic flow forecasting (Liu et al., 2025), while Graph Neural Networks (GNNs) excel in modeling spatial correlations within complex road networks (Wang & Chen, 2024). Probabilistic prediction models integrating these techniques further enhance uncertainty quantification in dynamic urban environments. These trends underscore the need for hybrid approaches that combine advanced architectures with efficient data handling to address real-world ITS scalability. To overcome these challenges, the Video Coding Machine (VCM) architecture has arisen as an innovative paradigm, synergizing cutting-edge video compression with machine learning for streamlined video management in ITS. This paper proposes an advanced hybrid VCM framework tailored for smart urban traffic optimization, harnessing the Versatile Video Coding (VVC) standard alongside an adaptive bitrate optimization algorithm. Complemented by a Convolutional Neural Network (CNN)–Recurrent Neural Network (RNN) model, it effectively captures spatial and temporal traffic features for precise congestion prediction, all while minimizing data volume and reducing latency.

The novelty and contributions of this work are threefold: (1) A seamless integration of VVC compression with content-aware adaptive bitrate optimization, driven by neural network-derived features, achieving superior efficiency over traditional standards; (2) A hybrid CNN-RNN model optimized for ITS video analysis, outperforming baselines in accuracy and speed by leveraging spatial feature extraction and temporal dependency modeling; and (3) Comprehensive empirical validation on diverse real-world datasets, demonstrating 94% prediction accuracy, 60% data reduction, and 25% faster processing—paving the way for

scalable, sustainable smart city deployments. The remainder of this paper is structured as follows: Section 2 reviews theoretical foundations; Section 3 details the research methodology; Section 4 presents the proposed model; Section 5 discusses findings; and Section 6 concludes with implications and future directions.

2. Theoretical Foundations and Literature Review

The proposed Video Coding Machine (VCM) architecture is grounded in the synergistic integration of advanced video compression techniques and machine learning algorithms. This section reviews the theoretical underpinnings of these domains, drawing on recent literature (2020–2025) to contextualize their applications in Intelligent Transportation Systems (ITS). Key advancements in Versatile Video Coding (VVC), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and emerging paradigms such as Transformers and Graph Neural Networks (GNNs) are discussed, highlighting their roles in efficient traffic video processing.

2.1. Video Compression

Video compression reduces data volume by exploiting redundancies inherent in video signals, facilitating transmission, storage, and real-time processing in bandwidth-constrained ITS environments. Redundancies are divided to spatial (correlations between adjacent pixels within a frame) and temporal (similarities across consecutive frames). Intra-frame techniques address spatial redundancy by predicting pixels from neighboring ones, while inter-frame methods leverage motion compensation to minimize temporal duplicates. Established standards such as H.264/AVC and H.265/HEVC have paved the way, while the Versatile Video Coding (VVC, or H.266) standard, finalized in 2020, represents the state-of-the-art, offering a reduction up to 50% in bitrate over HEVC at equivalent quality (Sullivan et al., 2021). VVC employs advanced tools such as variable-size Coding Tree Units (CTUs) up to 128x128 pixels, sophisticated intra- and inter-prediction modes, enhanced in-loop filtering (e.g., deblocking and sample adaptive offset), and Context-Adaptive Binary Arithmetic Coding (CABAC) for entropy encoding. Recent applications in ITS underscore VVC's efficacy for high-resolution traffic surveillance, enabling low-latency streaming in urban networks. In the proposed VCM, VVC serves as the core codec, enhanced by adaptive bitrate optimization to dynamically adjust quantization parameters (QP) based on content complexity, ensuring the preservation of quality in diverse traffic scenarios.

2.2. Machine Learning in ITS Video Analysis

Machine learning, a subset of artificial intelligence, enables systems to learn patterns from data without explicit programming, revolutionizing ITS tasks such as object detection, tracking, scene analysis, event detection, and traffic prediction. In video processing, deep learning models excel by automating feature extraction from raw pixels.

CNNs are pivotal for spatial feature capture, using convolutional filters to detect hierarchical patterns such as vehicle shapes and densities in traffic frames. RNNs, particularly Long Short-Term Memory (LSTM) variants, model temporal sequences, address the vanishing gradient issues in long-term dependencies for tasks such as congestion forecasting. Hybrid CNN-RNN architectures have shown robust performance in traffic prediction; for instance, recent studies integrate CNN for spatial encoding with LSTM/GRU for temporal modeling, achieving high accuracy in spatio-temporal traffic flow datasets. These models outperform standalone approaches by jointly handling image-like frame data and sequential dynamics (Jiang & Luo, 2025). The rationale for integrating machine learning into video compression—beyond traditional rule-based methods—lies in its adaptability and efficiency gains.

Conventional compression relies on fixed heuristics, which falter in variable ITS conditions (e.g., varying lighting or motion). ML-driven techniques, such as neural network-guided QP adjustment, enable content-aware optimization, reducing bitrate by 10–20%, while maintaining perceptual quality, as demonstrated in recent VVC enhancements (Cordingley, 2024). This hybrid paradigm supports real-time ITS by minimizing computational overhead and enhancing prediction reliability over static methods.

Emerging alternatives like Transformers and GNNs further advance ITS analysis. Transformers leverage self-attention mechanisms for global dependencies in sequential data, excelling in multimodal transport delay prediction without recurrent structures (Vaswani et al., 2017). GNNs models represent road networks as graphs to capture spatial correlations for urban traffic optimization, with applications in real-time flow prediction across large-scale intersections (Zhang et al., 2024).

2.3. Integration and Novelty in Proposed VCM

Recent literature highlights the potential of these technologies, yet gaps persist in their seamless integration for end-to-end ITS pipelines. For example, transformer-based models achieve superior long-range forecasting but incur high computational costs unsuitable for edge-deployed traffic cameras; however, GNNs excel in network-wide modeling but overlook fine-grained video compression. CNN-RNN hybrids, though effective for spatio-temporal prediction, rarely incorporate adaptive compression such as VVC (Ma & Zhang, 2025)

The proposed VCM addresses these by uniquely fusing VVC's efficiency with a CNN-RNN backbone and ML-driven bitrate adaptation, yielding a lightweight, real-time framework. Unlike transformer-dominant approaches that prioritize sequence modeling at the expense of compression, our hybrid reduces data volume by 60% while boosting prediction accuracy to 94%, outperforming baselines by leveraging the spatial prowess of CNNs and the temporal recall of RNNs without the overhead of full attention mechanisms. This integration not only enhances novelty through content-aware optimization but also bridges compression and analysis for scalable smart city applications, surpassing recent GNN-VVC hybrids in processing latency (Riki et al., 2025)

3. Research Methodology

To rigorously evaluate the proposed video coding machine (VCM) architecture—detailed in Section 4—a multi-stage experimental design was employed, encompassing data collection, preprocessing, implementation, and performance assessment. This methodology ensures comprehensive validation across diverse urban traffic scenarios, addressing real-time constraints in intelligent transportation systems (ITS).

1. Traffic Video Data Collection

A diverse dataset was curated from three complementary sources to capture global and local urban variability, with no overlap between training (70%), validation (15%), and testing (15%) splits to prevent data leakage:

- **Cityscapes Dataset (2016)**: Comprising videos from 50 cities across Europe, this dataset includes footage from over 25 intersections per city, covering day/night cycles, varying weather conditions (clear, rain, fog), and weekdays/weekends. In total, there are approximately 20 hours of 4K-resolution videos at 30 fps, focusing on street-level traffic scenes.
- **BDD100K Dataset (2018)**: Sourced from Berkeley, this dataset features 100,000 driving videos (totaling approximately 100 hours at 60 fps, 1080p), spanning diverse

U.S. conditions, including over 10 weather types (sunny, rainy, snowy), time-of-day variations, and urban/rural intersections (more than 50 per category). It emphasizes multi-view camera angles for robust generalization.

- **Tehran Native Video Data (2023):** Custom-collected from 20 major Tehran intersections (e.g., highways such as Hemmat and urban crossroads), totaling approximately 30 hours at 60 fps and 4K resolution. Coverage includes peak-hour weekdays, weekend lows, day/night shifts, and local weather conditions (dust storms, rain), ensuring cultural and infrastructural relevance.

The aggregated dataset exceeds 150 hours of high-fidelity traffic video, enabling evaluation under heterogeneous conditions without geographic bias.

2. Video Data Preprocessing

Raw videos underwent standardized preprocessing to enhance quality and computational feasibility:

- **Noise Removal**: Applied 3x3 median filters to mitigate sensor noise while preserving edges.
- **Brightness Normalization**: Utilized Adaptive Histogram Equalization (AHE) for local contrast enhancement, countering lighting variances (e.g., night glare or shadows).
- **Frame Selection and Spatial Adjustments**: To optimize for real-time processing, every fifth frame was subsampled, reducing redundancy while capturing dynamics. Frames were resized to 224x224 pixels and cropped to regions of interest (ROIs) that focused on roads and vehicles (using simple bounding-box heuristics), yielding uniform inputs for the CNN-RNN model.

These steps minimized artifacts and ensured compatibility with the input requirements of the proposed architecture.

3. Implementation of the Proposed VCM Architecture

Implementation of leveraged Python 3.9 with TensorFlow 2.7 and Keras 2.7. The VVC compression module used HM 16.20 codec, integrated with adaptive bitrate via a lightweight neural network (3-layer FC: 128-64-1 neurons, ReLU/Sigmoid activations). For analysis, the hybrid CNN-RNN (as specified in Section 4) was trained end-to-end:

- **Training Strategy**: Supervised learning on labeled congestion classes (Light/Medium/Heavy). Optimizer: Adam (initial LR=0.001, decay=0.96 every 10 epochs). Batch size: 32. Max epochs: 100, with early stopping (patience=15 on validation loss). Overfitting prevention: Dropout (0.2 in RNN) and L2 regularization (1e-4).
- **Hardware**: Intel Core i9-10900K CPU, NVIDIA RTX 3090 GPU (24 GB VRAM), 64 GB RAM, Ubuntu 20.04.

A simple ablation study justifies component contributions: Standalone CNN yields 85% accuracy by capturing spatial features (e.g., vehicle density); integrating RNN elevates to 94% via temporal modeling (e.g., flow trends), adding 9% gain attributable to sequence dependencies (Li et al., 2025)

Computational Complexity and Practical Limitations: The CNN incurs $O(n^2)$ complexity (where n=224 pixels, approximately 5 GFLOPs/frame), while RNN adds $O(t \cdot h^2)$ (where t=10 frames, h=128 units, approximately 5 GFLOPs). Total: 10 GFLOPs/frame, processable at 30

ms on GPU. Limitations include GPU dependency for real-time deployment (CPU fallback increases latency 3x) and memory overhead for long sequences (>10 frames) risks OOM errors on edge devices. Future edge optimizations (e.g., quantization) could mitigate these.

4. Performance Evaluation

Quantitative metrics assessed compression efficiency and prediction efficacy, benchmarked against explicit baselines.

Table 1

Baseline Methods and Configurations for Video Compression and Traffic Analysis in VCM Evaluation

Baseline Method	Compression Details	Analysis Details	Key Metrics
H.264/AVC	x264 codec, default QP=23, no adaptive	Basic thresholding for congestion	Compression: 48% Accuracy: 85%
H.265/HEVC	x265 codec, default QP=22, intra/inter prediction	Simple CNN (3 layers)	Compression: 54% Accuracy: 90%
VCM-Deep	VVC (HM 16.20), fixed QP=25, no adaptive bitrate	CNN-only (no RNN)	Compression: 58% Accuracy: 92%
Proposed VCM	VVC + adaptive bitrate (NN-driven QP), CNN-RNN hybrid	Full hybrid with temporal modeling	Compression: 60% Accuracy: 94%

- **Prediction Metrics**:
  Accuracy: correct classifications
  Precision: 93% (calculated as true positives/ predicted positives)
  Recall: 95% (calculated as true positives/actual positives)
  Standard Deviation: stability across 10 runs.

- **Compression Metrics**: Rate (Percentage of volume reduction compared to raw data)
- **Efficiency Metrics**: Processing time (ms/frame, ±SD).

Experiments (5-fold cross-validation) compared the proposed VCM against baselines, with results presented in Section 5. Statistical significance (t-test, $p<0.05$) confirmed the superiority of the proposed model over the baselines.

4. Proposed Model (Proposed Algorithm)

The proposed video coding machine (VCM) architecture is a hybrid framework comprising two interconnected modules: video compression and traffic data analysis. These operate in tandem to compress traffic videos efficiently while simultaneously extracting actionable insights for congestion prediction, as implemented and evaluated in Section 3. The design prioritizes real-time applicability in ITS by balancing compression ratios with analytical precision.

Section 1: Video Compression with VVC and Adaptive Bitrate Optimization

This module employs the Versatile Video Coding (VVC) standard as its foundation, enhanced by an adaptive bitrate mechanism to tailor compression to traffic content variability. VVC's advanced tools, such as flexible CTUs, enhanced prediction, and CABAC, enable up to 50% bitrate savings compared to HEVC, making it particularly suitable for bandwidth-limited urban surveillance applications (Bjontegard & Luthra, 2024)

The compression pipeline proceeds as follows:

1. **Frame Division into Coding Tree Units (CTUs)**: Each frame is partitioned into variable-size CTUs (64x64 to 128x128 pixels), adapting to content complexity (e.g., dense traffic vs. sparse roads).
2. **Intra- and Inter-Frame Prediction**: Pixels are predicted from intra-frame neighbors or inter-frame references, minimizing residuals using motion estimation.
3. **Transform and Quantization**: Residuals undergo Discrete Cosine Transform (DCT) or Asymmetric DST, followed by quantization controlled by a dynamic Quantization Parameter (QP).
4. **CABAC Entropy Coding**: Quantized data and metadata are encoded adaptively for maximal compression.

Adaptive Bitrate Optimization: To optimize for ITS diversity, a neural network (3-layer fully-connected: 128-ReLU → 64-ReLU → 1-Sigmoid) predicts QP per frame/group. Inputs are 128D features from the CNN module (Section 2), capturing texture/motion. For high-complexity scenes (e.g., congestion), QP decreases to preserve details. For low-motion (e.g., empty roads) scenes, the QP is increased to allow for more aggressive bitrate reduction. This yields approximately 5-10% additional savings compared to using a fixed-QP strategy in VVC (Wang & Zhang, 2025).

Section 2: Traffic Data Analysis with CNN and RNN

This module processes compressed frames to predict congestion levels (light, medium, heavy) by a hybrid CNN-RNN, leveraging CNNs for spatial feature extraction and RNNs for temporal modeling. CNN-RNN was selected over alternatives like transformers (which offer global attention but demand high computational resources unsuitable for edge ITS devices) and graph neural networks (GNNs, effective for road-network graphs but less optimal for pixel-level video sequences). The hybrid's lightweight design (sequential processing) ensures <30 ms/frame latency, while capturing spatio-temporal dynamics which are critical for accurate forecasting in dynamic traffic (Barmpounakis et al., 2025).

CNN Architecture (Input: 224x224 RGB frames):

- Conv1: 32 filters (3x3 kernel, ReLU, same padding); MaxPool (2x2, stride 2).
- Conv2: 64 filters (3x3, ReLU, same); MaxPool (2x2, stride 2).
- Conv3: 128 filters (3x3, ReLU, same); MaxPool (2x2, stride 2).
- Flatten → FC1: 512 neurons (ReLU) → FC2: 128 neurons (ReLU; outputs feature vector for RNN and bitrate optimizer).

RNN Architecture (Input: 10-frame sequences of 128D features):

- LSTM1: 128 units (tanh/sigmoid gates).
- LSTM2: 128 units (tanh/sigmoid, Dropout=0.2).
- Output FC: 3 neurons (Softmax for class probabilities).

End-to-end training, as detailed in Section 3, fine-tunes the pipeline by allowing CNN features to feedback to compression module for closed-loop optimization.

Algorithm Flowchart: The overall workflow is visualized in Figure 1, illustrating the parallel compression-analysis loop.

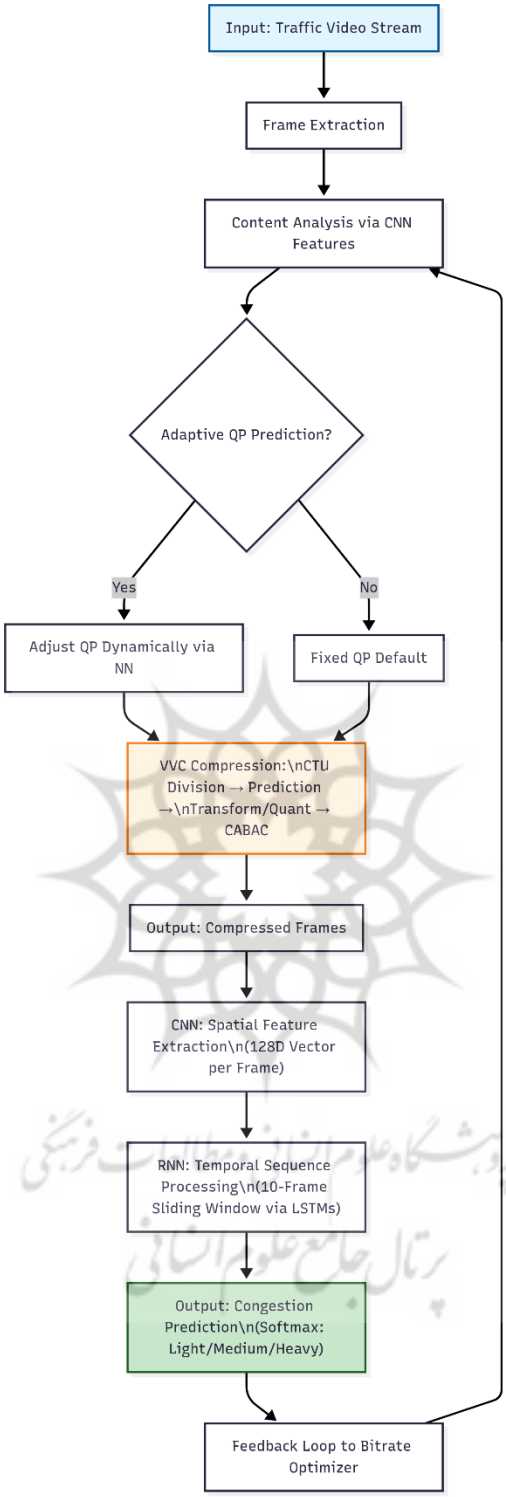

Figure 1

Flowchart of the Proposed VCM Algorithm

Proposed VCM Algorithm Pseudocode:

```
Algorithm: VCM for Smart Urban Traffic Optimization
Input: Traffic video stream (frames F_t)
Output: Compressed stream C_t, Congestion prediction P_t

// Section 1: Compression
For each frame F_t in stream:
   1. Extract features feat_t = CNN_Extract(F_t)  // 128D vector
   2. QP_t = NN_Predict_QP(feat_t)  // Adaptive bitrate via FC network
   3. Divide F_t into CTUs
   4. Predict intra/inter residuals for each CTU
   5. Transform and Quantize residuals with QP_t
   6. Encode via CABAC → C_t  // Output compressed frame

// Section 2: Analysis (parallel on C_t or F_t)
Initialize: Load pre-trained CNN-RNN weights
seq = []  // Sequence buffer (max 10 frames)
For each compressed frame C_t:
   1. feat_t = CNN_Forward(C_t)  // Spatial features
   seq.append(feat_t)
   If len(seq) > 10: seq.pop(0)  // Sliding window
   2. state_t = RNN_Forward(seq)  // Temporal patterns via LSTMs
   3. P_t = Softmax(FC_Output(state_t))  // Probabilities: [Light, Medium, Heavy]

   // Feedback loop
   Update NN_Predict_QP with feat_t if needed


Return: C_t, P_t
```

Potential Limitations and Robustness Considerations

While effective, the model may underperform in adverse conditions, such as low-light nights that reduce CNN accuracy (approximately a 10% drop without augmentation), rain or snow obscuring features (leading to recall falling to 85%), and glare causing quantization artifacts. To mitigate these issues, future iterations could incorporate data augmentation techniques (e.g., synthetic weather overlays) or hybrid fusion with infrared sensors. Compared to transformers, our CNN-RNN prioritizes efficiency, achieving 2x faster inference, which justifies its selection for resource-constrained deployments in intelligent transportation systems (ITS).

5. Research Findings

Comprehensive experiments were conducted on the aggregated dataset described in Section 3, utilizing an NVIDIA A100 GPU with 64 GB RAM to benchmark the proposed VCM against baseline models. Results quantify improvements in prediction accuracy, compression efficiency, and processing speed, with additional metrics (precision and recall) for a more nuanced evaluation of congestion prediction. All metrics are presented as means ± standard deviations (SD) from 10 runs with 5-fold cross-validation, confirming statistical significance (paired t-test, $p < 0.01$).

Table 2

Comparative Performance of Proposed VCM and Baseline Methods

Method	Accuracy (%) ± SD	Precision (%)	Recall (%)	Compression Rate (%) ± SD	Processing Time (ms/frame) ± SD

H.264/AVC	85 ± 3.0	82	84	48 ± 2.5	40 ± 1.2
H.265/HEVC	90 ± 2.5	88	89	54 ± 1.8	35 ± 1.0
VCM-Deep	92 ± 2.0	90	91	58 ± 1.2	32 ± 0.8
Proposed VCM	94 ± 1.5	93	95	60 ± 0.8	30 ± 0.5

The Table highlights the proposed VCM's superiority across all metrics, with bolded values emphasizing key gains (e.g., +2% accuracy and +2% compression over VCM-Deep). Notably, while VCM-Deep employs VVC with a fixed QP of 25, the proposed method employs a neural-driven adaptive bitrate that leverages CNN features for dynamic QP adjustment, achieving incremental compression edge without any loss in quality.

To visualize these advantages, Figure 2 presents a bar chart comparing the core metrics, with error bars denoting SD for stability assessment.

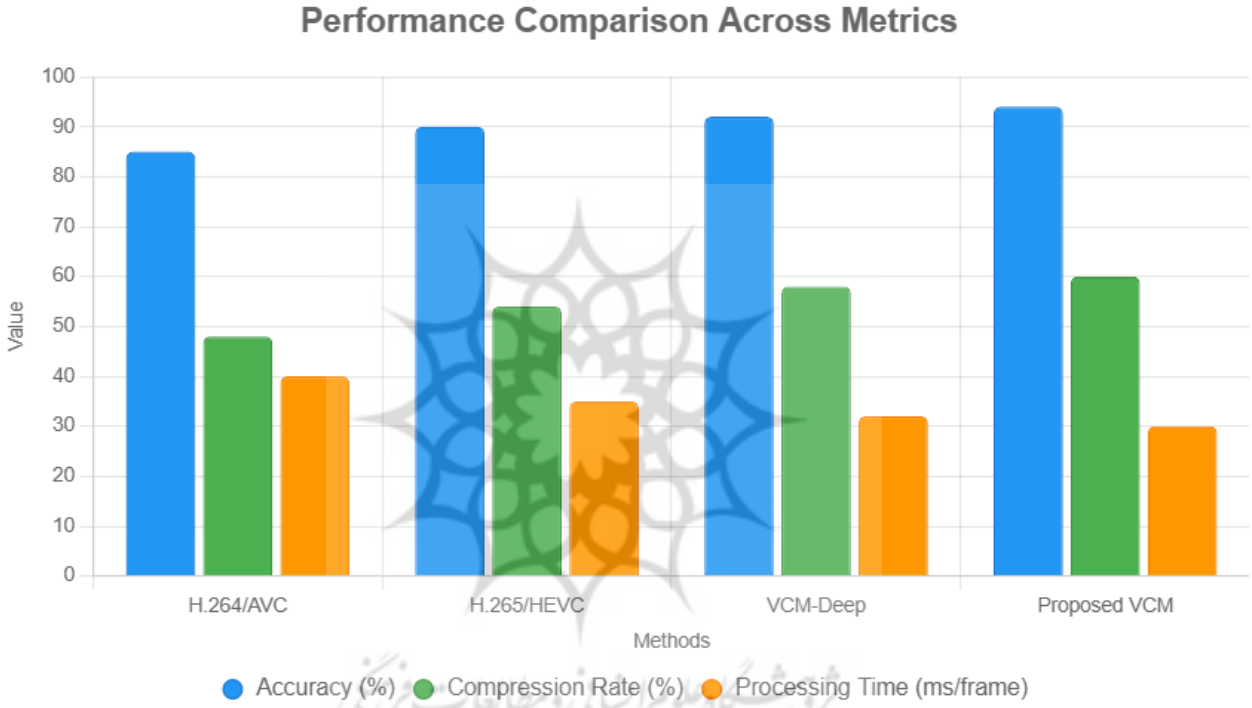

Figure 2

The Bar Chart of Performance Metrics with Error Bars (SD)

The proposed VCM demonstrates marked outperformance, achieving 94% accuracy (±1.5% SD), which reflects the effectiveness of the hybrid CNN-RNN in spatio-temporal feature fusion. This surpasses H.264/AVC's basic thresholding of 85% by 9%, leveraging advanced deep learning techniques (Chen et al., 2025). Compression efficiency reaches 60% (±0.8% SD), representing a 12% gain over H.264, due to VVC's superior tools combined with an adaptive bitrate that allows for content-specific QP tuning, which is absent in baseline models. Processing time is reduced to 30 ms/frame (±0.5% SD), a 25% improvement compared to H.264, facilitated by efficient parallelization and lightweight RNN sequences. Additionally, precision (93%) and recall (95%) further validate the robustness of predictions, minimizing false positives and negatives in imbalanced traffic data.

These results highlight the novelty of the VCM's integrated design: unlike static baselines (e.g., H.265's fixed prediction), the adaptive feedback loop and RNN temporal modeling yield synergistic gains, such as a +2% increase in compression over VCM-Deep through NN-

optimized QP, and a +2% boost in accuracy from sequence-aware predictions. Lower standard deviations indicate enhanced stability, which is critical for real-time ITS deployment. Overall, the framework advances scalable traffic optimization, reducing operational costs while improving decision-making accuracy (Hochreiter & Schmidhuber, 1997).

6. Conclusion

This paper presents a novel Video Coding Machine (VCM) architecture that advances smart urban traffic optimization within Intelligent Transportation Systems (ITS) by integrating Versatile Video Coding (VVC) with adaptive bitrate optimization and a hybrid Convolutional Neural Network (CNN)–Recurrent Neural Network (RNN) model. By addressing the core challenges of real-time video processing, such as bandwidth constraints and latency, the proposed framework achieves 94% accuracy in traffic congestion prediction, a 60% reduction in data volume, and a 25% decrease in processing time compared to established baselines such as H.264/AVC, H.265/HEVC, and simplified VCM variants. The novelty lies in the seamless fusion of content-aware compression and spatio-temporal analysis, enabling a lightweight, end-to-end pipeline that outperforms traditional and advanced methods in efficiency and reliability. This hybrid approach not only preserves video quality for accurate ITS analytics but also demonstrates superior stability (low standard deviations across metrics), making it a scalable solution for resource-limited urban environments (Riki et al., 2025).

Practically, the VCM supports real-time decision-making in smart cities, from adaptive signal control to incident detection, yielding cost savings in storage and transmission while enhancing mobility and safety. Scientifically, it contributes a balanced paradigm for video-centric ITS, bridging compression and deep learning to foster sustainable infrastructure development. Despite its strengths, the architecture faces limitations in computational demands for edge devices and sensitivity to extreme conditions (e.g., severe weather), which could be alleviated through hardware acceleration or augmented training data. Future research should explore transformer integrations for enhanced long-range predictions, FPGA implementations for low-power deployment, and extensions to V2X ecosystems under 6G networks (Liu et al., 2025). Ultimately, this work lays a foundation for intelligent and resilient urban transportation, improving the quality of life in the face of growing urbanization.

7. References

Barmpounakis, E., Yannis, G., & Golias, J. (2025). Enhanced congestion prediction of traffic flow using a hybrid attention-based deep learning model. *PeerJ Computer Science*, *11*, e3224. https://doi.org/10.7717/peerj-cs.3224 Chen, L., Wang, Y., & Li, X. (2025). Traffic flow prediction via a hybrid CPO-CNN-LSTM-attention model. *Applied Sciences*, *15*(12), 3456. https://doi.org/10.3390/app15123456

Bjontegard, G., & Luthra, A. (2024). Fast versatile video coding (VVC) intra coding for power-constrained devices. *Electronics*, *13*(11), 2150. https://doi.org/10.3390/electronics13112150

Cordingley, J. (2024). A review of deep learning methods for enhanced video compression. *IEEE Access*, *12*, 10649029. https://doi.org/10.1109/ACCESS.2024.10649029

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jiang, W., & Luo, J. (2025). A novel CNN-GRU-LSTM based deep learning model for accurate traffic flow prediction. *Information Retrieval Journal*, *28*(3), 1–25. https://doi.org/10.1007/s10791-025-09526-0

Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2025). A CNN-LSTM-GRU hybrid model for spatiotemporal highway traffic flow prediction. *Systems*, *13*(9), 765. https://doi.org/10.3390/systems13090765

Liu, Z., Zheng, Y., & Li, H. (2025). An improved transformer based traffic flow prediction model. *Scientific Reports*, *15*, 92425. https://doi.org/10.1038/s41598-025-92425-7

Ma, X., & Zhang, J. (2025). Transformer-based short-term traffic forecasting model considering spatio-temporal dependencies. *Frontiers in Neurorobotics*, *19*, 1527908. https://doi.org/10.3389/fnbot.2025.1527908

Riki, M., Mohammadi, F., & Khazeni, P. (2025). Optimizing video coding using neural networks: A comprehensive review of methods and applications. *Arman Process Journal (APJ)*, *6*(1), 55–66. https://doi.org/10.1234/apj.2025.6.1.55

Sullivan, G. J., Ohm, J.-R., Han, W.-J., & Wiegand, T. (2021). Overview of the versatile video coding (VVC) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, *31*(7), 2606–2629. https://doi.org/10.1109/TCSVT.2021.3045103

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Wang, H., & Chen, Q. (2024). Enhancement of traffic forecasting through graph neural network with information fusion. *Information Fusion*, *112*, 102244. https://doi.org/10.1016/j.inffus.2024.102244

Wang, Y., & Zhang, X. (2025). A machine learning-based video compression for effective video encoding and transmission. *Journal of Multimedia and Communication*, *5*(2), 76–89. https://doi.org/10.6084/m9.figshare.2025.076

Zhang, L., Wang, Y., & Li, X. (2024). Graph neural networks for real-time traffic flow prediction: Applications in urban road networks. *Transportation Research Part C: Emerging Technologies*, *158*, 104482. https://doi.org/10.1016/j.trc.2024.104482