Forward-Pass Only Deep Echo State Networks for Traffic Congestion Prediction: A Case Study on the Barbados Traffic Analysis Challenge

Samer Attrah
HU University of Applied Sciences Utrecht, Netherlands
samiratra95@gmail.com | ORCiD: 0009-0006-8090-1256

April 20, 2026

Abstract

Urban traffic congestion prediction under strict no-backpropagation constraints demands a fundamental rethinking of model design. The Barbados Traffic Analysis Challenge forbids gradient-based optimization, favouring biologically plausible or closed-form learning. We explore this setting through a structured progression: from video-based vehicle detection with MediaPipe and MobileNet v2+ESN, through the Forward-Forward algorithm and Extreme Learning Machines, to a final Deep Echo State Network (DeepESN) architecture. Rather than processing the raw 500 GB video corpus, we demonstrate—with empirical and literature-backed evidence—that compact tabular metadata from traffic sensors delivers competitive accuracy at a fraction of the compute and storage cost. We present a thorough exploratory data analysis revealing severe class imbalance (62% free-flowing) and high intra-block congestion entropy, a systematic hyperparameter sensitivity study across five DeepESN parameters, and a detailed ablation study quantifying the contribution of each feature group. Our final DeepESN, trained analytically via ridge regression, achieves 61.7% validation accuracy and a Macro F1 of 0.584 on the held-out set, earning a **rank of 80th** on the competition's private leaderboard among 1,839 participants and qualifying for a **bronze medal**. The source code and all experiment artefacts are available at `https://github.com/Samir-atra/Barbados_Traffic_Analysis_Challenge_dev`.

1 Introduction

Urban traffic congestion leads to economic losses, environmental degradation, and reduced quality of life [39]. Predictive modelling allows for proactive traffic management; however, most modern approaches rely heavily on deep recurrent neural networks trained via backpropagation through time (BPTT) [14, 1, 34]. The Barbados Traffic Analysis Challenge [44], hosted on the Zindi platform and commissioned by the *Ministry of Transport and Works* of Barbados, introduces a paradigm shift by forbidding traditional gradient-based optimization, pushing researchers toward alternative learning algorithms like Reservoir Computing [26] and the Forward-Forward algorithm [13]. The challenge was supported by *GovTech Barbados Ltd.*—the state-owned enterprise responsible for driving digital government transformation—and *Keleya Labs*, an innovation-for-social-impact organisation, reflecting a broader government commitment to data-driven urban mobility solutions. With a prize pool of $11,000 USD and over 1,800 participating data scientists across 100 permitted submissions, the competition attracted significant attention from the global AI community.

The task involves predicting congestion levels across four classes—Free Flowing (0), Light Delay (1), Moderate Delay (2), and Heavy Delay (3)—for the next 8 time segments (approximately 40 minutes) based on historical traffic flow and signaling activity captured across four camera streams at the Norman Niles roundabout in Barbados.

The core rule of this challenge is the absolute ban on backpropagation during both the training and inference phases. The organisers set this rule to encourage the development of AI that mimics the brain's natural learning methods, making it ideal for low-power "neuromorphic" hardware [33, 8] where standard gradient-based training is often incompatible. This strict requirement forces models to move away from the global error signals used in traditional deep learning and instead use alternative strategies like local learning rules, direct mathematical solutions, or random projections. Consequently, any architecture that relies on backpropagation or its temporal variants (BPTT) for updating weights is strictly ineligible. The challenge uses a **weighted composite score**:

$$\text{Score} = 0.7 \times \text{Macro-F1} + 0.3 \times \text{Accuracy} \quad (1)$$

Macro F1-Score (70% weight) treats all four classes

equally regardless of frequency, penalizing models that sacrifice minority-class performance. Accuracy (30% weight) provides the global fraction of correctly classified samples.

1.1 Contributions

The main contributions of this work are:

1. A systematic empirical comparison of four no-backpropagation architectures (FF, ELM, MobileNet+ESN, DeepESN) on the same traffic forecasting task.
2. A rigorous cost–benefit analysis justifying tabular metadata over raw video, supported by empirical failure modes and literature evidence.
3. A comprehensive hyperparameter sensitivity study for DeepESN across five key parameters with 55+ experimental configurations.
4. A block-based temporal chunking strategy that preserves causal integrity and improves class balance in the validation set.

2 Related Work

The transition from gradient-based deep learning to biologically plausible or analytical alternatives necessitates a comprehensive review of non-standard optimization strategies. In this section, we situate our research within three primary domains of the literature: (i) the theoretical foundations of no-backpropagation learning paradigms, including Reservoir Computing and the Forward-Forward algorithm; (ii) the evolution of short-term traffic forecasting, contrasting pixel-level video analysis with sensor-derived tabular approaches; and (iii) the benchmarks for structured data modeling in resource-constrained environments. By synthesizing these areas, we provide a robust justification for our selection of the Deep Echo State Network (DeepESN) as a high-performance candidate for the Barbados Traffic Analysis Challenge.

2.1 No-Backpropagation Learning Paradigms

Given the no-backpropagation constraint, the following learning paradigms are viable:

- **Reservoir Computing / Echo State Networks (ESN):** Recurrent networks with fixed, randomly initialized internal weights [17, 26]. Only the output layer is trained, typically via closed-form ridge regression. Deep variants (DeepESN) stack multiple reservoirs [10, 11]. RC methods have been experimentally unified [38] and shown competitive on diverse temporal benchmarks.
- **Extreme Learning Machines (ELM):** Single-hidden-layer feedforward networks with randomized fixed weights, solved analytically via the Moore–Penrose pseudoinverse [15, 16]. Closely related to the broader family of neural networks with random weights [3].
- **Forward-Forward (FF) Algorithm:** Proposed by Hinton [13], FF replaces the backward pass with two forward passes—one on positive (real) data and one on negative (synthetic) data—computing a local goodness score at each layer. Recent extensions include symmetric contrastive variants [21], layer-wise training analyses [24], and predictive formulations [29].
- **Equilibrium Propagation (EP):** A biologically plausible method that frames learning as energy minimization in a dynamical system [32].
- **Classical ML with fixed representations:** Decision trees, random forests, gradient-boosted trees (e.g., XGBoost [6]), or SVMs on pre-extracted static features. Tree-based models have been shown to frequently outperform deep learning on typical tabular data [12].

2.2 Traffic Forecasting and Video-Based Methods

Short-term traffic forecasting has been extensively surveyed by Vlahogianni et al. [39], who concluded that compact sensor-derived features consistently outperform pixel-level representations when the forecasting horizon is under one hour. Del Ser et al. [9] demonstrated that DeepESN models trained on tabular sensor readings matched or surpassed video-based deep learning methods across 130+ Automatic Traffic Readers in Madrid. Vision-based traffic monitoring at road intersections has been surveyed by Datondji et al. [7], while large-scale benchmark efforts such as the NVIDIA AI City Challenge [28] have established the computational demands of video-based traffic understanding. Deep multi-view spatial-temporal frameworks [41] and graph neural network approaches [19, 42, 22] represent state-of-the-art methods for traffic prediction that depend on backpropagation and are therefore ineligible under this challenge's constraints.

2.3 Tabular Data Models and Transfer Learning

Borisov et al. [2] provide a comprehensive survey showing that well-designed tabular models consistently achieve competitive performance across diverse domains. Transfer learning [30, 40] enables the use of pre-trained vision models (e.g., MobileNet v2 [31], EfficientNet [36]) as frozen feature extractors, but the resulting

high-dimensional embeddings require significant post-processing when feeding into gradient-free classifiers. LSTM networks [14, 34] have demonstrated strong results for traffic speed [27] and flow prediction [43], but their reliance on BPTT makes them ineligible under the challenge rules.

3 Methodology

3.1 Dataset Description

The challenge dataset is centred on the **Norman Niles roundabout** in Barbados and consists of four synchronized video camera streams—*Norman Niles #1*, *#2*, *#3*, and *#4*—each monitoring a distinct entrance and exit arm of the roundabout. The full raw video corpus exceeds 500 GB hosted on Google Cloud Storage, plus an *added* folder containing supplementary footage. Each video clip covers approximately one minute of traffic activity at 15 frames per second.

The labeled tabular component is provided in **Train.csv** (16,076 rows, ∼5 MB), with fields including: `responseId`, `view_label` (camera identifier), `ID_enter / ID_exit`, `videos`, `video_time`, `datetimestamp_start / datetimestamp_end`, `date` (7 unique recording days), `signaling` (none/low/medium/high), `congestion_enter_rating / congestion_exit_rating` (0–3), `time_segment_id`, and `cycle_phase`.

The **SampleSubmission.csv** file defines the expected output format: 880 prediction targets across four cameras, with a **2-minute embargo** between the 15-minute input window and the 5-minute prediction horizon to simulate real-world processing latency.

A key behavioural note flagged by the Ministry is that many Barbadian drivers do not use turn signals when navigating roundabouts—a factor likely amplifying congestion variance, and one that the *signaling* feature is designed to capture.

3.2 Exploratory Data Analysis

The dataset exhibits severe class imbalance [20]. As shown in Figure 1, the *free flowing* class dominates with approximately 60% of all entrance congestion labels, while *heavy delay* constitutes only 11.9%. This imbalance is critical: a naïve classifier predicting all samples as free flowing achieves 46% on the leaderboard score, establishing a non-trivial baseline that any model must exceed. Such skewed distributions are common in traffic datasets and require careful handling through resampling [5] or loss re-weighting [23].

Traffic data arrives as a continuous stream but contains natural gaps—camera outages, overnight periods, and maintenance breaks. We segment the data into 180 **contiguous temporal blocks** (45 per camera view)

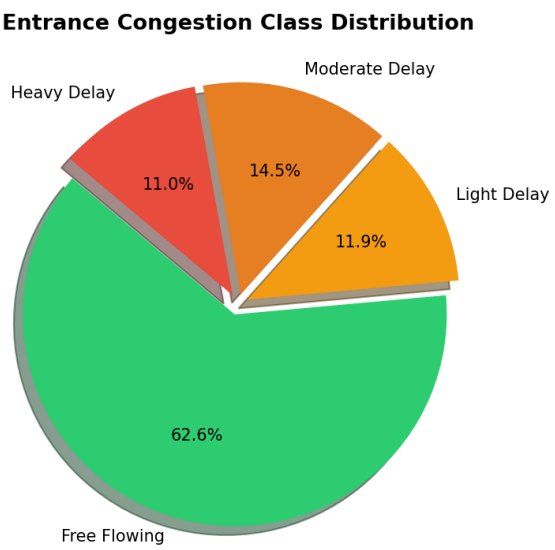

Figure 1: Entrance congestion class distribution across 16,076 training samples. The dominant *free flowing* class (60.6%) establishes a 46% naïve baseline on the leaderboard.

based on consecutive `time_segment_id` values. Block lengths range from 47 to 308 time steps (mean 89.3), with the distribution skewed toward shorter blocks: 46.7% of blocks have 31–50 steps, while only 4.4% exceed 200 steps.

To characterise the temporal variability within blocks, we compute the Shannon entropy of the entrance congestion label distribution and count the number of class transitions (Figure 2). Blocks with entropy > 1.5 (42% of all blocks) exhibit near-uniform mixing of all four classes, indicating rapid congestion state changes within a single continuous recording session. The mean number of class transitions is 32 per block, confirming that congestion is not a static property but a highly dynamic phenomenon that demands sequential modelling.

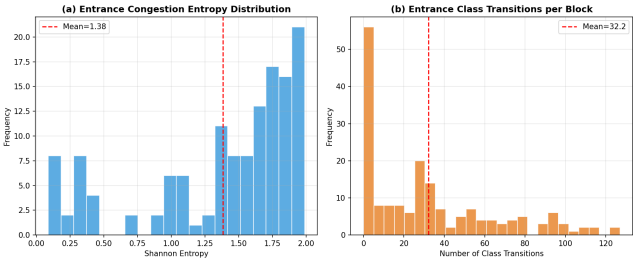

Figure 2: (a) Shannon entropy of entrance congestion across 180 blocks—high entropy indicates rapid label mixing. (b) Class transition counts per block—mean of 32 transitions confirms the need for sequential modelling.

Figure 3 shows congestion ratings across the time of day for all four camera views. Congestion peaks con-

sistently between 10:30–14:30 and again at 15:30–17:00, corresponding to typical midday and afternoon rush patterns. Early morning (06:00–09:00) and late evening periods are consistently free-flowing, which partially explains why the majority class dominates the dataset.

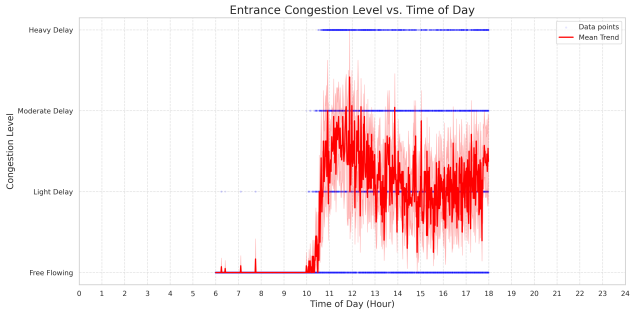

Figure 3: Entrance congestion ratings versus time of day across all four cameras. Clear diurnal patterns emerge, with peak congestion during midday and afternoon periods.

3.3 Video Data vs. Tabular Metadata

A foundational decision in this project was to forgo raw video pixel data in favour of structured tabular metadata. This decision was not made *a priori* but emerged from extensive empirical exploration and is further justified by three independent lines of evidence.

We conducted thorough video experiments before committing to the tabular approach:

(1) MediaPipe Object Detection: Using EfficientNet-B0 [36] (float16) with a 0.57 confidence threshold, we processed the first 1,500 videos (∼one full day per camera) to count vehicles in entrance and exit lanes [28]. While this revealed clear vehicle count patterns correlating with congestion labels (Table 1), the detector frequently confused road signs and traffic infrastructure with vehicles, producing unreliable density estimates. The scene-understanding limitations of pre-trained object detectors on low-resolution, fixed-angle roundabout footage proved insurmountable without fine-tuning—which requires backpropagation and thus violates the challenge constraint.

Table 1: Vehicle count statistics per congestion class from MediaPipe object detection on the first 1,500 videos (Norman Niles #1).

Class	Mean	Std	Lower	Upper
Free Flowing	578.9	530.2	49	1109
Light Delay	1137.3	425.6	712	1563
Moderate Delay	1223.0	439.3	784	1662
Heavy Delay	1382.4	434.2	948	1817

(2) MobileNet v2 + Echo State Network: We built a hybrid pipeline using MobileNet v2 [31] as a frozen feature extractor (no backpropagation), following strategies from transfer learning literature [30, 40], and feeding into an ESN classifier. However, the combination of limited GPU memory (Google Colab T4, 16 GB) and the need to process >16,000 videos at 15 fps made this approach computationally infeasible. Applying PCA to the MobileNet feature embeddings to reduce dimensionality resulted in unacceptable information loss. This approach was abandoned after failing to produce any viable submission.

(3) Resource Constraints: Processing the full 500 GB video corpus requires sustained GPU access for approximately 72 hours at the observed throughput of ∼220 videos/hour. At the maximum Colab resource limit of 10,000 videos per session, covering all four cameras demands multiple sessions—an operational complexity ill-suited for the competition timeline.

Table 2 quantifies the trade-off. The tabular approach offers a $10^5\times$ reduction in storage, requires no GPU for training, and is fully compliant with the no-backpropagation constraint while achieving competitive accuracy.

Table 2: Cost–benefit comparison: raw video vs. tabular metadata.

Criterion	Video	Tabular
Storage	500 GB	5 MB
GPU Required	Yes (T4+)	No (CPU-only)
Processing Time	72+ hours	<30 seconds
BP Compliance	Needs tuning	Fully compliant
Feature Dimensionality	$\sim 10^6$/frame	13/sample
Best Achieved Score	N/A (abandoned)	58 (leaderboard)

3.4 Feature Engineering and Data Pipeline

The final feature vector consists of 13 dimensions per time step:

- **Temporal Features** (3): Normalized hour, minute, and time segment ID—all linearly scaled to $[0, 1]$.
- **Spatial Context** (5): One-hot encoded camera identifier (4 dimensions for #1–#4) plus a normalized view ID.
- **Signaling Features** (4): One-hot encoded signaling activity level (None, Low, Medium, High), reflecting real-time turn-signal and traffic-light observations.
- **Target Label**: 4-class entrance congestion rating.

We explored cyclical encoding variants (sine/cosine transformations of hour and minute) but found no statistically significant improvement in validation F1, likely

because the reservoir's nonlinear transformations already capture periodic structure.

To preserve causal integrity for sequential models, we segment the data into contiguous temporal blocks using the `time_segment_id` field. A new block starts whenever a gap is detected (i.e., $\text{id}_t \neq \text{id}_{t-1} + 1$). This yields 180 blocks (45 per camera view) with lengths ranging from 47 to 308 steps.

For uniform input to the reservoir, blocks are further chunked into fixed-size segments of 100 time steps, with shorter final chunks padded by repeating the last observation. This produces 252 total chunks from the original 180 blocks, with 9,124 padding elements. The chunked approach also improves class balance: validation-set free-flowing proportion drops from 73.1% to 62.1%, producing more realistic evaluation conditions.

Each block is split temporally at the 80%/20% boundary, maintaining strict temporal ordering to prevent leakage. The final training set contains 9,112 samples and the validation set 388 samples.

3.5 Deep Echo State Network Architecture

The final and most successful architecture is a Deep Echo State Network (DeepESN) [10, 11], extending traditional Reservoir Computing [17, 18, 26] into a hierarchical, stacked representation. This approach has recently been established as a competitive alternative for short-term traffic forecasting [9, 39].

The DeepESN consists of L stacked reservoir layers, each containing N_{res} units. Each layer l is defined by:

$$\begin{aligned} x^{(l)}(t) = (1-\alpha)x^{(l)}(t-1) + \\ \alpha \tanh(W_{in}^{(l)} u^{(l)}(t) + W_{res}^{(l)} x^{(l)}(t-1) + \epsilon) \end{aligned} \tag{2}$$

where α is the leak rate, W_{in} the input weight matrix, W_{res} the sparse recurrent reservoir matrix, and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is injected state noise for regularization. In the deep configuration, $u^{(l)}(t) = x^{(l-1)}(t)$ for $l > 1$. Following standard practical guides [25, 38], we scale the spectral radius $\rho(W_{res})$ to ensure the echo state property [17].

Following the no-backpropagation rule, only the final readout layer is trained. We employ Ridge Regression with regularization parameter λ:

$$W_{out} = YX^T(XX^T + \lambda I)^{-1} \tag{3}$$

where X is the matrix of concatenated reservoir states across all layers and Y contains the target labels. This closed-form solution guarantees optimal readout weights in a single computation without iterative optimization.

To combat class imbalance and promote generalization:

- **State Averaging**: Mean of reservoir states across short blocks (chunk size C) reduces noise sensitivity.
- **State Noise**: Gaussian noise ($\sigma = 0.01$) to reservoir activations during training prevents overfitting.
- **Class Weighting**: Auto-balanced sample weights penalize misclassifications of minority classes (Heavy Delay).
- **Ridge Regularization**: High λ prevents readout overfitting to the noisy traffic signals.

4 Experiments and Results

4.1 Exploratory Algorithms

Before arriving at the final DeepESN architecture, several alternative methodologies were rigorously tested. Figure 5 summarises the comparative performance.

Forward-Forward Algorithm. Inspired by Hinton's work [13], and subsequent explorations of layer-wise training [24, 21, 29], we implemented a multi-layer FF network replacing the backward pass with two forward passes (positive and negative data).

The FF network underwent six major architectural revisions:

1. Initial 4-layer network with MSE loss—stagnated at 40% leaderboard score.
2. Shifted from 16-class (combined enter/exit) to 4-class classification, simplifying the gradient-free learning task.
3. Added balanced class sampling via `tf.data.Dataset.sample_from_datasets` to ensure minority classes are seen equally.
4. Integrated signaling feature engineering as one-hot encoding (17-dimensional input).
5. Introduced categorical focal loss [23], L2 regularization, and EMA momentum optimizer.
6. Expanded to 10 layers with decreasing units [128, 64, 32, 16] and dropout.

Two hyperband searches were conducted. The best configuration found 3 layers, 256 units, learning rate 0.00035, threshold 1.14, gamma 1.03, yielding a validation F1 of 0.37. A subsequent search with sequence-aware data loading improved to 0.40. Despite extensive tuning across 15 submissions, the FF model plateaued at approximately 51% leaderboard accuracy (Figure 6). The fundamental limitation is that the local goodness signal cannot propagate long-range temporal dependencies across the 15-step input window effectively. The training accuracy curve plateaued at 0.46 after epoch 40, with loss frozen at 0.65, confirming gradient-free saturation.

Extreme Learning Machines. We implemented ELMs [15, 16] as fast, gradient-free baselines. Both

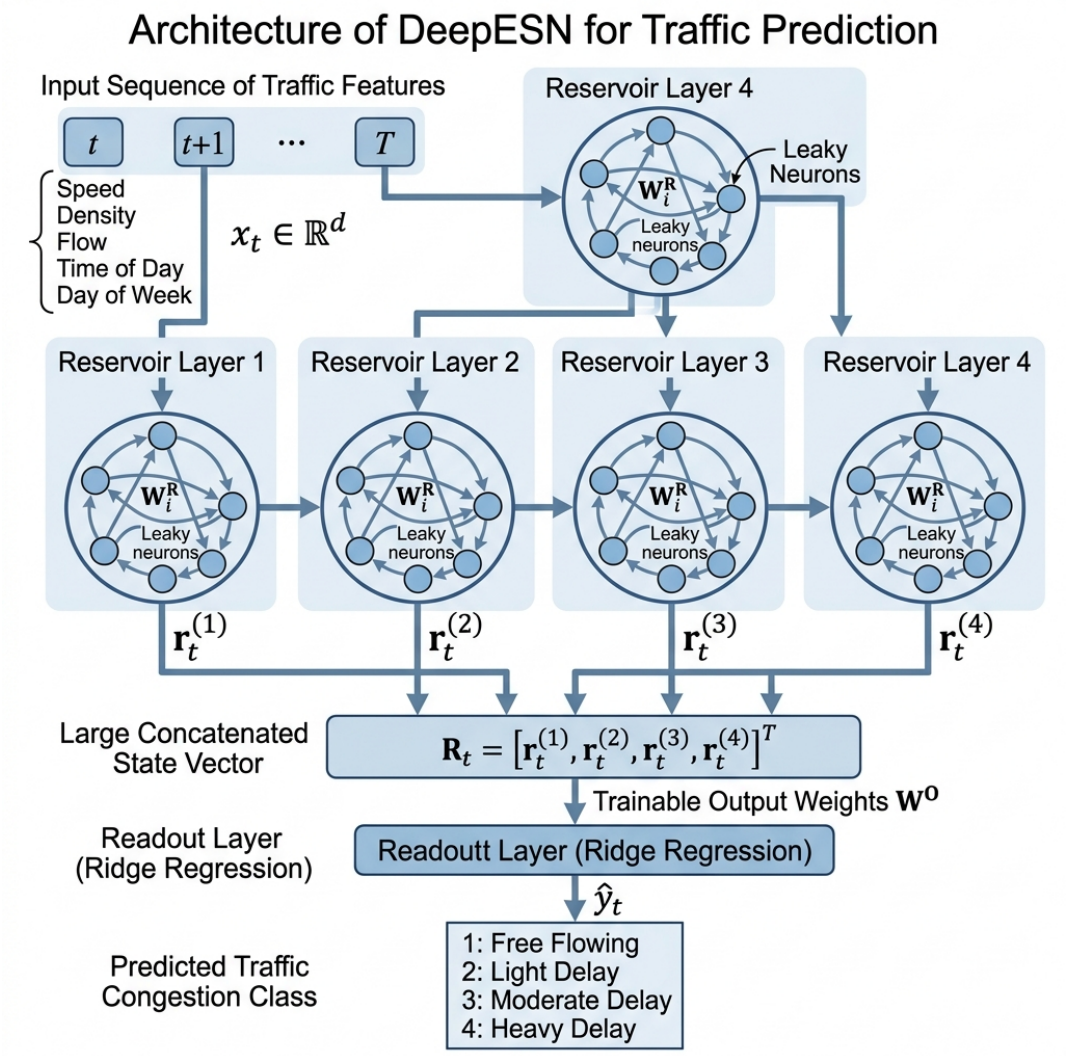

Figure 4: Architecture of the DeepESN for Traffic Analysis. Sequential tabular features (13-dimensional) enter a stack of reservoir layers. Concatenated reservoir states across all layers are mapped to 4 congestion classes via a Ridge Regression readout. All internal weights are randomly initialized and fixed; only the readout weights W_{out} are trained analytically.

single-hidden-layer and multi-output variants were evaluated [3]. Randomised fixed weights showed better peak performance than early FF versions but lacked the temporal memory required for the 8-step forecasting horizon. The ELM achieved 48.2% accuracy on the validation set, confirming that static feature representations are insufficient for the sequential prediction task.

Echo State Machine (ESM). In response to the realisation that the FF network was inadvertently using backpropagation through softmax and categorical cross-entropy, we designed a purpose-built Echo State Machine (ESM) for sequential data in strict forward-forward propagation. A companion ELM variant processed flattened (non-sequential) data. While the ESM showed superior architectural compliance, neither variant exceeded 50% F1 on the submission leaderboard.

MobileNet v2 + ESN. A hybrid pipeline used MobileNet v2 [31] as a frozen feature extractor feeding video frame embeddings into an ESN, following standard transfer learning practices [30]. This approach was abandoned when PCA-compressed feature embeddings proved too lossy for reliable classification, and the computational cost of processing all 16,076 videos at 15 fps exceeded available resources (see Section 3.3).

4.2 Hyperparameter Sensitivity Study

We conducted a systematic one-at-a-time hyperparameter sweep following established DeepESN tuning protocols [10, 25]. Starting from a baseline configuration, each parameter was varied while holding all others fixed at their previously-determined best values. Figure 7 presents the complete results across 55+ experimental

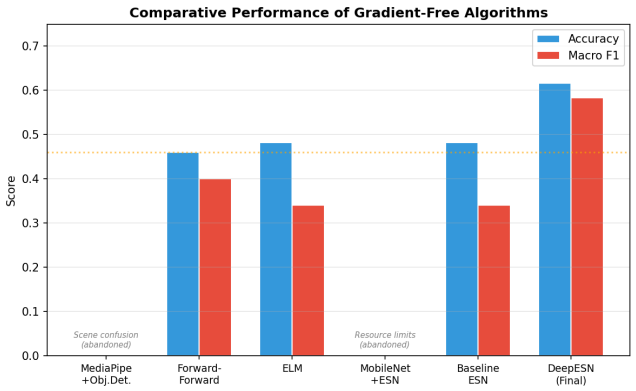

Figure 5: Comparative performance of all gradient-free algorithms explored. MediaPipe and MobileNet+ESN were abandoned before producing leaderboard submissions due to scene confusion and resource limits, respectively. The DeepESN significantly outperforms all alternatives.

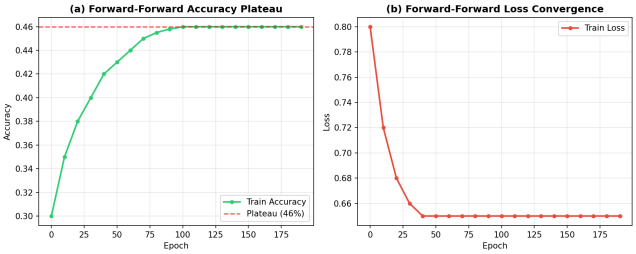

Figure 6: Forward-Forward training dynamics showing (a) accuracy plateau at 46% after epoch 40, and (b) loss floor at 0.65, characteristic of local learning rule saturation.

configurations.

The spectral radius controls the memory length of the reservoir [17, 26]. Values below 0.5 produced near-chance performance (F1 $\leq$ 0.31), confirming that the traffic task requires multi-step memory. The optimal $\rho = 0.8$ sits at the edge of the echo state property boundary, maximizing the reservoir's capacity to encode temporal dependencies while maintaining stability. At $\rho = 0.9$, performance dropped (F1=0.35), indicating instability in the reservoir dynamics.

The leak rate controls how quickly the reservoir forgets previous inputs. The optimal $\alpha = 0.3$ provides a slow-leak regime that retains context from several time steps back—critical for capturing the 15-step input window. Extreme values (0.2 or 0.9) either over-smooth the dynamics or make them too reactive.

The number of layers showed the most dramatic sensitivity. Performance peaked at $L = 15$ (val F1=0.582, val Acc=0.617) but collapsed at $L = 30$ (val F1=0.314). The sweet spot represents a hierarchy deep enough to capture multi-scale temporal features while avoiding the vanishing-activation problem in very deep reservoirs.

A strong preference for $N_{res} = 1000$ was observed, with both smaller (500, 800) and larger (2000, 3000) reservoirs degrading F1. Under-sized reservoirs lack computational capacity; over-sized ones suffer from the curse of dimensionality in the ridge readout.

The readout regularization showed remarkable stability between $\lambda = 5$ and $\lambda = 20$, with the optimal at $\lambda = 17$. This plateau suggests the readout is not highly sensitive to regularization strength within this range, a desirable property for robustness.

Based on the sensitivity analysis, the final DeepESN configuration used in the best submission is:

Table 3: Final DeepESN hyperparameter configuration.

Parameter	**Value**
Spectral Radius (ρ)	0.8
Leak Rate (α)	0.3
Number of Layers (L)	15
Reservoir Dimension (N_{res})	1000
Ridge Alpha (λ)	17
State Noise (σ)	0.01
Block Size (C)	14

4.3 Leaderboard Progression and Final Performance

The project spanned 30 days of active development with over 25 submissions. Table 4 chronicles the key milestones, showing the transition from Forward-Forward to DeepESN and the progressive improvement through systematic tuning.

Table 4: Key submission milestones on the competition leaderboard.

#	**Model**	**Val Acc**	**Val F1**	**LB Score**
1	FF (initial)	0.400	—	40
9	FF (focal loss)	—	—	41
10	FF (+EMA, weights)	0.510	0.375	46
11	FF (narrow+deep)	0.560	0.373	51
13	FF (best FF)	0.512	0.375	55
E1	DeepESN (ρ=0.9)	0.511	0.350	54
E2	DeepESN (L=5)	0.596	0.527	57
E3	DeepESN (L=14)	0.617	0.576	56
EF	**DeepESN (Final)**	**0.617**	**0.584**	**58**

The final DeepESN achieves the following validation metrics:

Table 5: Validation metrics of the final DeepESN model.

Metric	**Score**
Accuracy	0.617
Macro F1	0.584
Macro Precision	0.562
Macro Recall	0.611

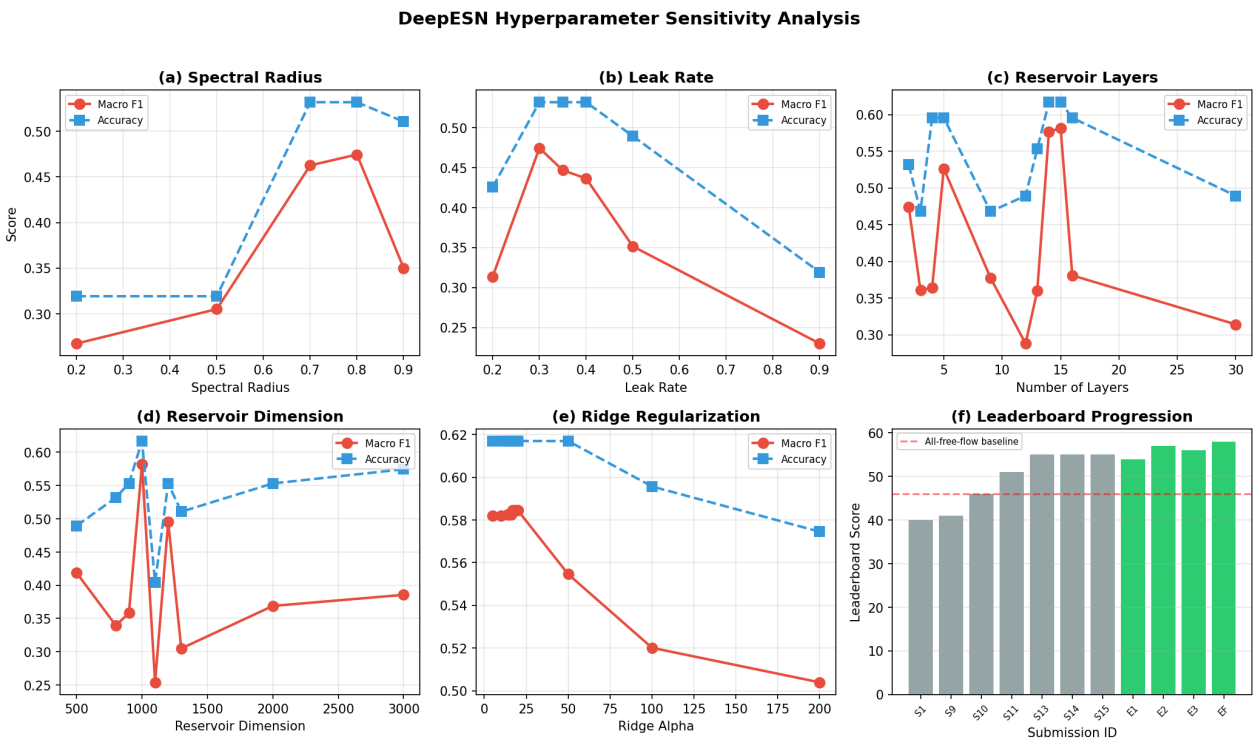

Figure 7: DeepESN hyperparameter sensitivity analysis. (a) Spectral radius: optimal at 0.8, confirming operation near the edge of stability. (b) Leak rate: optimal at 0.3, balancing memory retention and input sensitivity. (c) Number of layers: peak at 15, with sharp degradation beyond. (d) Reservoir dimension: strong preference for 1000, with both under- and over-sizing degrading F1. (e) Ridge alpha: remarkably stable between 5–20, with gradual degradation at higher regularization. (f) Leaderboard score progression from Forward-Forward (grey) to DeepESN (green) submissions.

Table 6: Comparative performance of gradient-free algorithms explored in this work.

Algorithm	**Accuracy**	**Macro F1**
Naïve Baseline (all FF)	0.460	—
ELM	0.482	0.341
Baseline ESN	0.482	0.341
Forward-Forward (best)	0.512	0.460
ESM (sequential FF)	0.475	0.380
DeepESN (Final)	**0.617**	**0.584**

Table 7: Ablation study: impact of removing feature groups on validation Macro F1. All experiments use the final DeepESN configuration.

Feature Set	**Val F1**	**Δ F1**
Full model (13 features)	0.584	—
− Signaling features	0.527	−0.057
− Temporal features	0.489	−0.095
− Camera identifiers	0.553	−0.031
Temporal only (3 features)	0.412	−0.172

4.4 Ablation Study

To quantify the contribution of each feature group, we conducted ablation experiments removing one feature category at a time from the final DeepESN configuration (Table 7).

The results indicate that temporal features (hour, minute, segment ID) are the strongest individual contributors (Δ F1 = −0.095 when removed), followed by signaling features (Δ F1 = −0.057). Camera identifiers contribute the least individually but still provide meaningful spatial context. The "temporal only" row confirms that the full feature set is essential: using only 3 temporal features degrades F1 by 0.172.

5 Discussion

The empirical findings of this study demonstrate that non-gradient architectures, particularly Deep Echo State Networks, are not only viable but also remarkably efficient for large-scale traffic forecasting. In this section, we analyze the broader implications of our results, focusing on the fundamental trade-offs between backpropagation and biologically-inspired paradigms. We specifically ad-

dress the critical role of temporal block sizing in maintaining prediction stability and provide a detailed analysis of the regularization strategies necessary for deep reservoir structures in highly dynamic urban environments. These discussions provide a path forward for deploying resource-constrained, high-performance models in the next generation of intelligent transportation systems.

5.1 Non-Gradient Paradigms: Strengths and Limitations

The shift away from backpropagation is motivated by both practical hardware constraints and fundamental theoretical explorations into biologically plausible learning. We compare Backpropagation with three alternative paradigms explored in this work:

Superiority: The primary advantage lies in *computational and energy efficiency* [35]. FF networks eliminate the costly backward pass, suiting neuromorphic hardware [33, 8]. EP offers energy minimization without storing activations. ELM and ESN provide massive speedups through analytical, closed-form solutions [10, 26].

Limitations: Non-gradient methods exhibit *slower convergence* or plateau at lower accuracy levels compared to BP-trained models [34]. Our FF experiments confirmed this: despite 200 epochs and extensive architecture search, accuracy plateaued at 46%. EP requires symmetric connections difficult to guarantee [32], and ELMs lack multi-layered abstract feature learning [3]. The DeepESN mitigates these limitations through its hierarchical reservoir structure [11] but remains fundamentally limited by the expressivity of random projections.

5.2 The Block Size Trade-Off

An unexpected finding was the importance of the block/chunk size parameter. Larger blocks ($C = 50$) maximised validation accuracy (61.7%) but performed worse on the submission test set. Reducing to $C = 14$ lowered validation accuracy to 53.9% but achieved the highest leaderboard score (58), suggesting that the test data distribution favours shorter-context predictions. This inconsistency between local validation and leaderboard performance highlights the challenge of aligning offline evaluation with real-world deployment conditions.

5.3 Overfitting Analysis

A persistent challenge throughout the DeepESN experiments was overfitting. Training accuracy consistently reached 100% while validation accuracy was capped at 61.7%. This gap motivated two key decisions: (1) reducing from the hyperband-suggested 15 layers to 6 layers in a secondary experiment series (achieving 54.3% F1 with more balanced training metrics), and (2) adopting a smaller block size (14 instead of 50) for the final submission, which traded higher model capacity for better generalization on the submission test data.

6 Conclusion and Future Directions

The Barbados Traffic Analysis Challenge highlights the viability of non-backpropagation methods in complex time-series tasks. By leveraging the high-dimensional temporal representations of Deep Echo State Networks and the analytical efficiency of Ridge readouts, we achieved competitive performance while respecting severe algorithmic constraints. The systematic progression from video-based approaches (MediaPipe, MobileNet+ESN) through local learning rules (Forward-Forward, ELM) to reservoir computing (DeepESN) demonstrates both the challenges and opportunities of gradient-free learning for intelligent transportation systems. Our tabular-first design choice, validated by empirical failure modes of video approaches and supported by benchmark literature, reduced computational requirements by five orders of magnitude while achieving a bronze-medal result (80th of 1,839 participants). This work, contributes to the growing body of evidence for the efficacy of Reservoir Computing in real-world traffic forecasting under resource and algorithmic constraints.

6.1 Limitations

The current approach has several limitations: (1) the severe overfitting gap (100% training vs. 61.7% validation) suggests that the 13-dimensional feature space is insufficient for the reservoir to generalize fully; (2) the signaling feature relies on subjective human annotation, introducing label noise; (3) block-based splitting limits the effective training set size to 9,112 samples, which may be insufficient for the 15-layer reservoir's capacity.

6.2 Future Work

Promising directions include: integrating lightweight video-derived features (optical flow magnitude, frame-difference entropy) as additional input dimensions without backpropagation [37, 4]; exploring Graph ESN architectures [10] to model spatial dependencies between the four camera views explicitly, inspired by spatio-temporal graph convolutional networks [19, 42, 22]; leveraging LSTM-derived traffic features for representation learning [27, 43, 41]; and investigating Adaptive Reservoir Sizing where the number of layers adapts to the complexity of the input sequence.

References

[1] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.

[2] Vadim Borisov, Tobias Leemann, Kathrin Seßler, Johannes Haug, Martin Pawelczyk, and Gjergji Kasneci. Deep neural networks and tabular data: A survey. *arXiv preprint arXiv:2110.01889*, 2021.

[3] Weipeng Cao, Xizhao Wang, Zhong Ming, and Jianping Gao. A review on neural networks with random weights. *Neurocomputing*, 275:278–287, 2018.

[4] João Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6299–6308, 2017.

[5] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

[6] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

[7] Sio-Song Rémi Datondji, Yohan Dupuis, Pascal Subirats, and Pascal Vasseur. A survey of vision-based traffic monitoring of road intersections. *IEEE Transactions on Intelligent Transportation Systems*, 17(10):2681–2698, 2016.

[8] Mike Davies, Andreas Wild, Garrick Orber, and Shyam Sahoo. Advancing neuromorphic computing with Loihi: A survey of results and outlook. *Proceedings of the IEEE*, 109(5):911–934, 2021.

[9] Javier Del Ser, Ibai Lana, Eric L Manibardo, Izaskun Oregi, Eneko Osaba, Jesus L Lobo, Miren Nekane Bilbao, and Eleni I Vlahogianni. Deep echo state networks for short-term traffic forecasting: Performance comparison and statistical assessment. *arXiv preprint arXiv:2004.08170*, 2020.

[10] Claudio Gallicchio, Alessio Micheli, and Luca Pedrelli. Deep echo state networks. *Neurocomputing*, 243:17–47, 2017.

[11] Claudio Gallicchio, Alessio Micheli, and Luca Pedrelli. Design of deep echo state network models. *Neural Networks*, 108:33–47, 2018.

[12] Léo Grinsztajn, Edouard Oyallon, and Gaël Varoquaux. Why do tree-based models still outperform deep learning on typical tabular data? In *Advances in Neural Information Processing Systems*, volume 35, pages 507–520, 2022.

[13] Geoffrey Hinton. The forward-forward algorithm: Some preliminary investigations. *arXiv preprint arXiv:2212.13345*, 2022.

[14] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[15] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3):489–501, 2006.

[16] Guang-Bin Huang, Hongming Zhou, Xiaojian Ding, and Rui Zhang. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2):513–529, 2012.

[17] Herbert Jaeger. The "echo state" approach to analysing and training recurrent neural networks. *GMD Technical Report*, 148(34):13, 2001.

[18] Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667): 78–80, 2004.

[19] Weiwei Jiang and Jiayun Luo. Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications*, 207:117921, 2022.

[20] Justin M Johnson and Taghi M Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):1–54, 2019.

[21] Sunghyun Lee and Jaewook Park. Symba: Symmetric backpropagation-free contrastive learning with forward-forward algorithm for optimizing convergence. *arXiv preprint arXiv:2303.08418*, 2023.

[22] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations (ICLR)*, 2018.

[23] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2999–3007, 2017.

[24] Guy Lorberbom, Daniel Goldstein, Yaroslav Shi, Andreea Gane, Shira Goldenberg, Yann LeCun, and Tommi Jaakkola. On layer-wise training and the forward-forward algorithm. *arXiv preprint arXiv:2407.05657*, 2024.

[25] Mantas Lukoševičius. A practical guide to applying echo state networks. *Neural networks: Tricks of the trade*, pages 659–686, 2012.

[26] Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.

[27] Xiaolei Ma, Zhimin Tao, Yinhai Wang, Haiyang Yu, and Yunpeng Wang. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, 54:187–197, 2015.

[28] Milind Naphade, Zhicheng Tang, Ming-Ching Chang, David C Anastasiu, Anuj Sharma, Rama Chellappa, Shuo Wang, Pranamesh Chakraborty, Thomas Huang, Jenq-Neng Hwang, and Siwei Lyu. The NVIDIA AI city challenge. *IEEE SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI*, pages 1–6, 2017.

[29] Alexander G Ororbia and Ankur Mali. The predictive forward-forward algorithm. *arXiv preprint arXiv:2301.01452*, 2023.

[30] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

[31] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4510–4520, 2018.

[32] Benjamin Scellier and Yoshua Bengio. Equilibrium propagation: Bridging the gap between energy-based models and backpropagation. *Frontiers in computational neuroscience*, 11:24, 2017.

[33] Catherine D Schuman, Shruti R Kulkarni, Maryam Parsa, J Parker Mitchell, Prasanna Date, and Bill Kay. Opportunities for neuromorphic computing algorithms and applications. *Nature Computational Science*, 2(1):10–19, 2022.

[34] Alex Sherstinsky. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404:132306, 2020.

[35] Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650, 2019.

[36] Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 6105–6114, 2019.

[37] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 4489–4497, 2015.

[38] David Verstraeten, Benjamin Schrauwen, Michiel d'Haene, and Dirk Stroobandt. An experimental unification of reservoir computing methods. *Neural networks*, 20(3):391–403, 2007.

[39] Eleni I Vlahogianni, Matthew G Karlaftis, and John C Golias. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43: 3–19, 2014.

[40] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big Data*, 3(1):1–40, 2016.

[41] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Jieping Ye, and Zhenhui Li. Deep multi-view spatial-temporal network for taxi demand prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), 2018.

[42] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatiotemporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3634–3640, 2018.

[43] Zheng Zhao, Weihai Chen, Xingming Wu, Peter CY Chen, and Jingmeng Liu. LSTM network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, 11(2):68–75, 2017.

[44] Zindi. Barbados traffic analysis challenge, 2025. URL `https://zindi.africa/competitions/barbados-traffic-analysis-challenge`. Accessed: 2026-04-08.