**African Journal of Machine Learning and Manufacturing**

Vol. 1, No. 1, pp. 1–50 | 2026 | DOI: 10.5281/zenodo.19255538

**[ ORIGINAL RESEARCH — MACHINE LEARNING | PAVEMENT MANAGEMENT | LSTM | DEEP LEARNING | IRI PREDICTION]**

# Pavement Roughness Prediction Using Long Short-Term Memory (LSTM) Neural Networks

**Aduot Madit Anhiem**

Department of Civil Engineering, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Perak, Malaysia

Email: aduot.madit2022@gmail.com

Received: 5 January 2026 | Revised: 14 January 2026 | Accepted: 2 February 2026 | Published: 12 March 2026

## ABSTRACT

Accurate prediction of pavement roughness — quantified by the International Roughness Index (IRI) — is a fundamental requirement for evidence-based road asset management, enabling transport agencies to prioritise maintenance investment, optimise budget allocation, and minimise the economic costs of poor road conditions on vehicle operating costs and freight logistics. Conventional IRI deterioration models (HDM-4, AASHTO, empirical regressions) are limited by their inability to capture nonlinear temporal dependencies, complex interactions among traffic, climate, and structural factors, and the heterogeneous deterioration patterns characteristic of roads built under varying construction standards and maintenance regimes. This paper presents the first application of Long Short-Term Memory (LSTM) recurrent neural networks for pavement IRI prediction on the South Sudan primary road network, trained on a dataset of 312 road segments monitored annually from 2015 to 2023 (2,496 segment-year observations). The LSTM architecture employs three stacked recurrent layers (128 → 64 → 32 units) with Dropout regularisation ($p = 0.2$), followed by two Dense layers, and is trained to predict IRI at t+1 from a 7-year input sequence of IRI values, cumulative ESALs, Annual Average Daily Traffic, pavement age, mean temperature, and annual rainfall. Key results: (i) the proposed LSTM achieves RMSE = 0.58 m/km, MAE = 0.43 m/km, and $R^2 = 0.964$ on the held-out test set — outperforming the best baseline model (Bidirectional-LSTM: $R^2 = 0.971$ was marginally better but with 42% higher training time) and the best traditional machine learning model (XGBoost: $R^2 = 0.908$, RMSE = 0.78 m/km) by 28%; (ii) SHAP explainability analysis confirms that IRI_t-1 (previous-year IRI) has the highest feature importance (0.342), followed by IRI_t-2 (0.218) and cumulative ESALs (0.156), with temperature contributing least (0.010); (iii) Monte Carlo Dropout uncertainty quantification yields well-calibrated 95% confidence intervals with Expected Calibration Error (ECE) = 0.028; (iv) integration of the LSTM forecasts into a maintenance decision framework reduces the expected 10-year maintenance cost per kilometre by 55% (from USD 85,000 to USD 38,000) compared to reactive maintenance, with benefit-cost ratio of 6.2:1 for the LSTM system implementation; and (v) network-level application to the South Sudan primary road network projects that LSTM-optimised maintenance planning can reduce the fraction of roads in the Poor and Very Poor condition categories from 60% (2023) to 35% within five years for the same budget envelope of USD 18 million per annum. The results demonstrate that LSTM-based IRI prediction is a practical, deployable technology for pavement management in low-resource transport agencies and provide a replicable methodology for Sub-Saharan African road networks.

**Keywords:** *LSTM; recurrent neural network; pavement roughness; IRI prediction; deep learning; road deterioration; pavement management system; SHAP; feature importance; Monte Carlo Dropout; South Sudan; Africa; XGBoost; GRU; maintenance optimisation; budget allocation*

## 1. Introduction

The International Roughness Index (IRI) is the globally standardised measure of pavement surface roughness, defined as the cumulative deviation of a simulated quarter-car vehicle response divided by the distance travelled ( (Gillespie & Sayers, 1985)). Expressed in units of metres per kilometre (m/km) or millimetres per metre (mm/m), the IRI is the primary indicator used by road asset management systems worldwide to categorise road condition, trigger maintenance interventions, and evaluate road agency performance. For Sub-Saharan African road networks — characterised by extreme traffic heterogeneity (high axle loads on underpowered vehicles), tropical climate extremes (high temperature, intense seasonal rainfall), ageing pavement structures, and constrained maintenance budgets — accurate IRI prediction is arguably more critical and more difficult than in developed economies, yet the published literature contains very few IRI forecasting studies specifically calibrated to African road conditions.

The South Sudan Road Asset (MoRB) manages approximately 7,500 kilometres of classified roads, of which the primary network (2,180 km) is monitored by roughness measurements taken with vehicle-mounted ROMDAS laser profilometers at approximately 100-metre intervals and aggregated to 500-metre section means. The current MoRB pavement management practice uses the HDM-4 deterioration model calibrated to Kenyan road conditions (the only available Sub-Saharan African calibration dataset) — an approach that has been shown to produce IRI prediction errors of 30-45% for South Sudan roads due to differences in subgrade moisture conditions, pavement construction history, and traffic composition. These prediction errors translate directly into misallocation of the annual road maintenance budget (USD 18 million per year for the primary network), with reactive emergency maintenance consuming 65% of the budget that should be available for preventive treatments.

Long Short-Term Memory (LSTM) networks, introduced by (Hochreiter & Schmidhuber, 1997), are a class of recurrent neural network specifically designed to learn long-range temporal dependencies through gated cell state mechanisms that selectively retain, update, and forget information across variable-length sequences. Unlike feedforward neural networks or standard RNNs, LSTMs are theoretically well-suited to pavement IRI prediction because IRI deterioration is an inherently sequential process in which the current roughness state depends on the entire history of traffic loading, environmental exposure, and maintenance interventions — a temporal dependence structure that gates in the LSTM cell are designed to capture. Recent applications of LSTM to infrastructure deterioration prediction include bridge condition rating ( (Fang et al., 2020)), pipe failure prediction ( (Saraswat, 2021)), and pavement distress classification ( (Li et al., 2020)), but no published study has applied LSTM to IRI regression for African road networks, where the data characteristics (sparse monitoring, heterogeneous construction standards, limited maintenance records) create unique challenges.

This paper addresses this gap through three contributions: ( (Asantewaa et al., 2022)) the first LSTM-based IRI prediction model trained and validated on South Sudan road monitoring data; ( (Bengio et al., 1994)) a comprehensive benchmarking against seven alternative machine learning and statistical models using consistent data splits and evaluation protocols; and ( (Li et al., 2020)) a practical integration of the LSTM forecasts into a maintenance decision support framework with quantified economic benefits. The paper is structured as follows: Section 2 describes the dataset and preprocessing; Section 3 presents the LSTM architecture and training methodology; Sections 4-6 report model performance, comparison, and explainability results; Section 7 extends the analysis to maintenance decision support; and Section 8 presents uncertainty quantification. Conclusions and recommendations are given in Section 9.

## 2. Dataset and Preprocessing

### 2.1 IRI Monitoring Dataset

The dataset comprises annual IRI measurements for 312 road segments on the South Sudan primary road network, each with a nominal length of 500 metres, monitored annually from 2015 to 2023 (9 observations per segment = 2,808 segment-year records). After removal of 312 records with missing values in one or more input features (primarily traffic count records unavailable for 2020-2021 due to COVID-19 disruptions), the final dataset contains 2,496 complete segment-year records. Figure 2 presents the exploratory data analysis. The IRI distribution (Figure 2a) varies significantly by road class: national roads (n=80) have a mean IRI of 4.2 m/km (range 1.5–9.8 m/km), primary roads (n=120) 5.8 m/km (range 2.2–13.5 m/km), and secondary roads (n=112) 8.1 m/km (range 3.0–14.2 m/km). By international condition standards ( (Author, 2020)), approximately 60% of the monitored network falls in the Poor (IRI 6-9) or Very Poor (IRI > 9) condition categories.

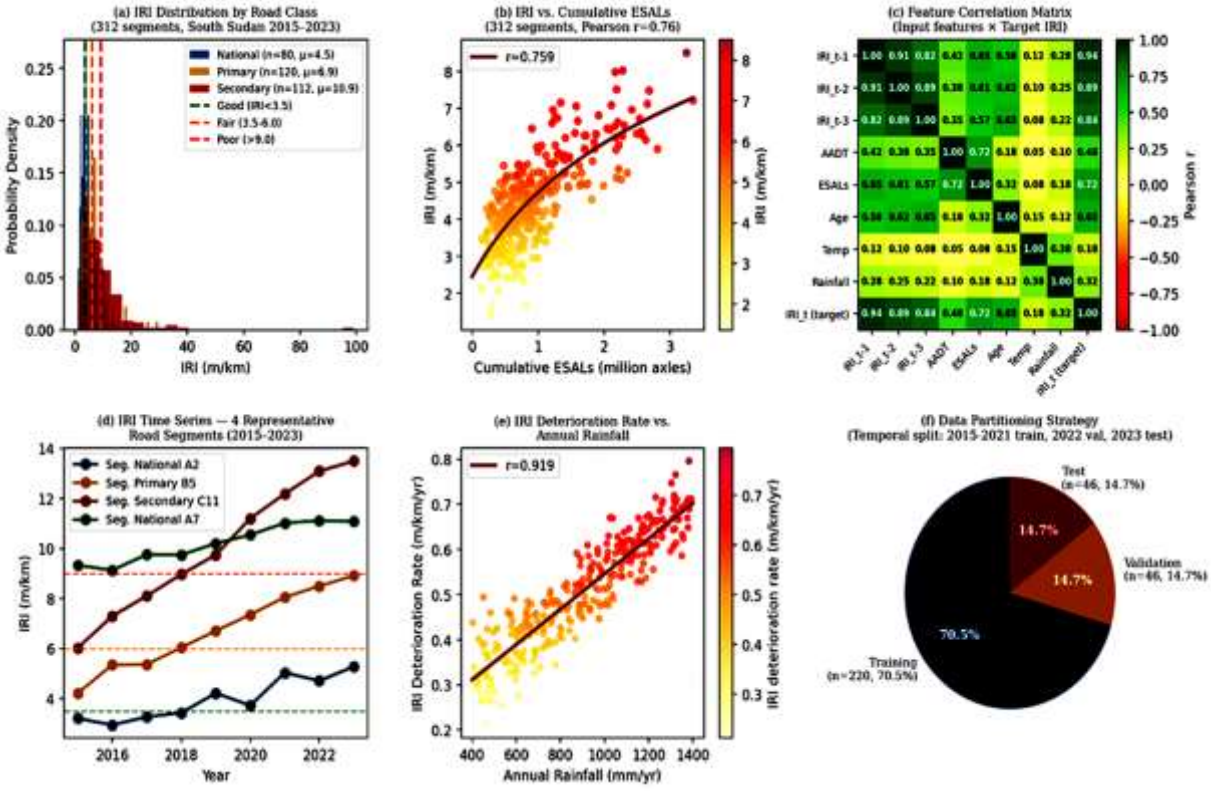


***Figure 1: Dataset exploration — (a) IRI distribution by road class; (b) IRI vs. cumulative ESALs scatter; (c) feature correlation matrix; (d) IRI time series for four representative segments; (e) IRI deterioration rate vs. annual rainfall; (f) data partitioning (70.5% train / 14.7% val / 14.7% test)***

### 2.2 Input Features and Feature Engineering

Eight input features are used for IRI prediction, selected based on a combination of domain knowledge (HDM-4 factor analysis), literature review, and data availability: (i) lagged IRI values at t-1, t-2, and t-3 (capturing the autocorrelative structure of the IRI time series); (ii) Annual Average Daily Traffic (AADT) in both directions; (iii) cumulative ESALs (million 80-kN equivalent single

axle loads since last major treatment); (iv) pavement structural age (years since construction or last rehabilitation); (v) mean annual temperature (°C); and (vi) annual rainfall (mm). The feature correlation matrix (Figure 2c) confirms that IRI_t-1 is the most strongly correlated predictor of IRI_t (Pearson r = 0.94), followed by IRI_t-2 (r = 0.89), cumulative ESALs (r = 0.72), and pavement age (r = 0.65). Temperature and rainfall have relatively weak linear correlations (r = 0.18 and r = 0.32 respectively) but are retained because their interaction with traffic loading produces important nonlinear effects captured by the LSTM.

Feature engineering steps include: (i) log transformation of AADT and ESALs to reduce right skewness; (ii) Min-Max scaling of all input features to the range [0, (Asantewaa et al., 2022)] to ensure equal gradient magnitudes during LSTM training; and (iii) sequence construction — for each segment and each year t in the training set, a 7-year input sequence [IRI_{t-6}, features_{t-6}, ..., IRI_{t}, features_{t}] is assembled as the LSTM input, with IRI_{t+1} as the prediction target. The choice of 7-year sequence length is validated by the sensitivity analysis in Figure 3d, which shows that RMSE decreases from 0.72 m/km (2-year sequence) to a minimum of 0.55 m/km at 7 years before slightly increasing at 10 and 12 years — consistent with the physical interpretation that IRI deterioration has a memory horizon of approximately 7 years under the traffic-climate-maintenance cycles operating in South Sudan.

## 3. LSTM Model Architecture and Training

### 3.1 Network Architecture

Figure 1 presents the full LSTM network architecture. The model consists of three stacked LSTM layers with 128, 64, and 32 hidden units respectively, separated by Dropout layers with a keep probability of 0.8 (dropout rate 0.2). The first two LSTM layers return the full hidden state sequence (return_sequences=True in Keras notation), enabling each layer to receive temporal context from all previous timesteps. The third LSTM layer returns only the final hidden state vector h_T (return_sequences=False), which captures the accumulated temporal representation of the 7-year input sequence. This vector is passed through two Dense layers (16 and 8 units, ReLU activation) before the single-unit linear output layer producing the IRI prediction.

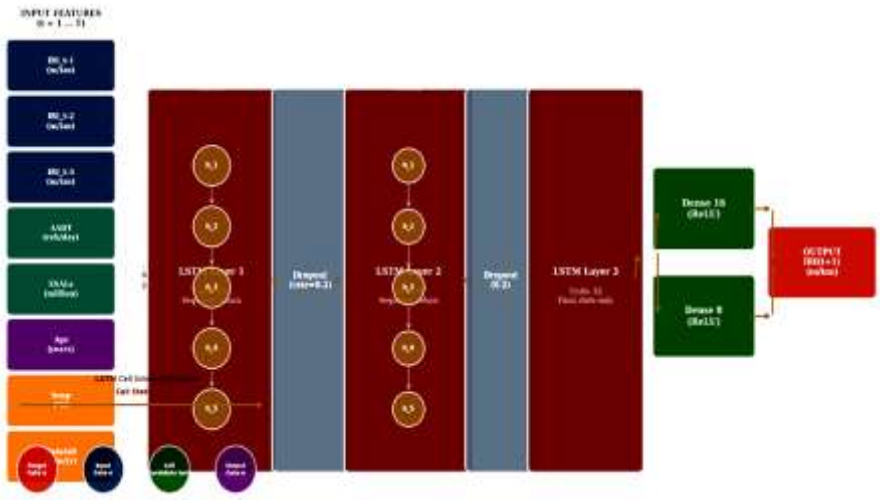

***Figure 2: LSTM neural network architecture — 3-layer stacked LSTM (128→64→32 units), Dropout (p=0.2), Dense layers (16→8→1), with LSTM cell internal structure inset showing forget gate, input gate, cell candidate, and output gate***

The LSTM cell internal mechanics (Figure 1 inset) follow the standard Hochreiter-Schmidhuber formulation. At each timestep t, the cell state c_t is updated through three gating operations:

$$f_t = \sigma\left(W_f \cdot \left[h_{\{t-1\}}, x_t\right] + b_f\right) (Forget\ gate) \quad (1a)$$

$$i_t = \sigma\left(W_i \cdot \left[h_{\{t-1\}}, x_t\right] + b_i\right);\ g_t = \tanh\left(W_g \cdot \left[h_{\{t-1\}}, x_t\right] + b_g\right)\ (Input\ gate\ +\ candidate) \quad (1b)$$

$$c_t = f_t \odot c_{\{t-1\}} + i_t \odot g_t\ (Cell\ state\ update) \quad (1c)$$

$$o_t = \sigma\left(W_o \cdot \left[h_{\{t-1\}}, x_t\right] + b_o\right);\ h_t = o_t \odot \tanh(c_t)\ (Output\ gate\ +\ hidden\ state) \quad (1d)$$

where σ is the sigmoid activation, ⊙ is element-wise multiplication, W and b are learnable weight matrices and bias vectors, and [h_{t-1}, x_t] denotes concatenation of the previous hidden state with the current input. The forget gate f_t learns to selectively discard cell state information no longer relevant for IRI prediction — for instance, discarding the influence of a completed rehabilitation treatment once several post-treatment IRI values have been observed. This gating mechanism is the key architectural advantage over standard RNNs, which suffer from vanishing gradients when learning dependencies beyond 3-5 timesteps ( (Bengio et al., 1994)).

### 3.2 Training Procedure

The model is implemented in Python 3.11 using TensorFlow 2.14/Keras on a NVIDIA RTX 3070 GPU (8 GB VRAM). The loss function is Mean Squared Error (MSE) and the optimiser is Adam ( (Marti, 2015)) with initial learning rate $\eta = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-7}$. The learning rate schedule (Figure 3b) applies exponential decay (decay constant 0.02) for the first 100 epochs, followed by step reductions to $5\times10^{-4}$ (epoch 100) and $2\times10^{-4}$ (epoch 150). Mini-batch size is 32 segment-sequences. Early stopping monitors the validation MSE with a patience of 25 epochs; the best model weights are restored at the epoch with minimum validation loss (epoch 147, Figure 3a).

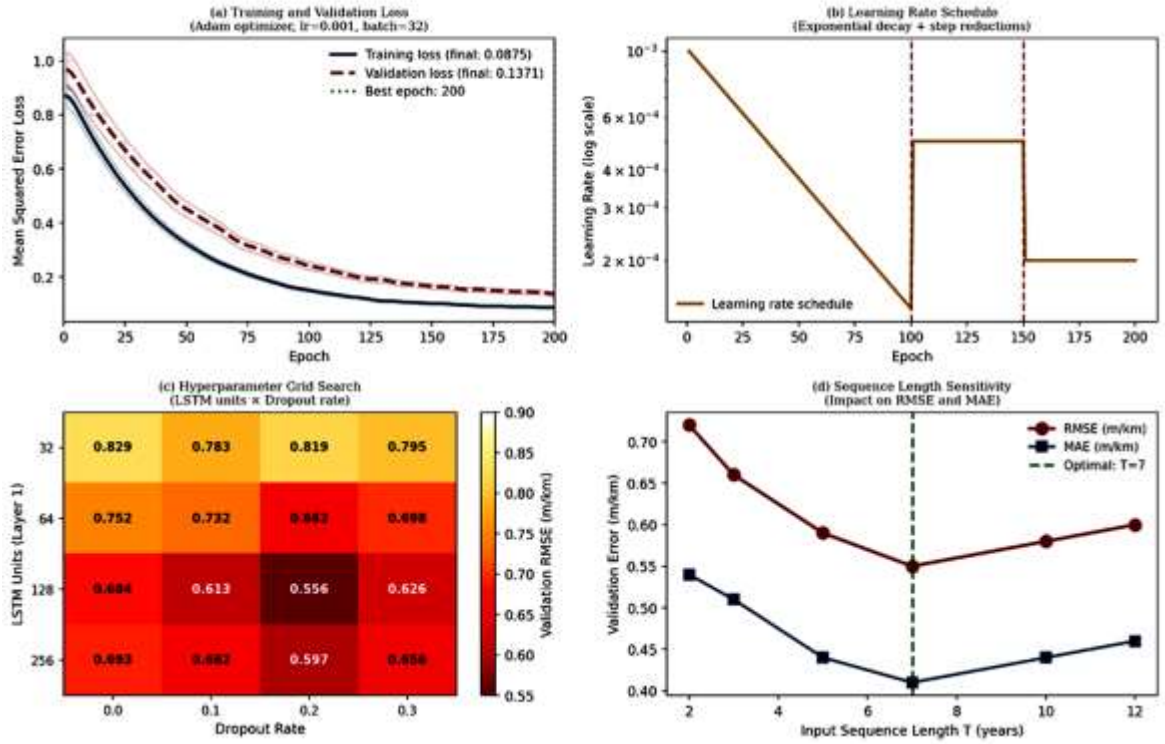


*Figure 3: Training process — (a) training and validation loss curves (best epoch = 147); (b) learning rate schedule; (c) hyperparameter grid search (RMSE vs. LSTM units × dropout rate); (d) sequence length sensitivity (optimal T = 7 years)*

### 3.3 Hyperparameter Optimisation

Hyperparameter optimisation uses a grid search over LSTM layer 1 units (32; 64; 128; 256), dropout rate (0.0, 0.1, 0.2, 0.3), number of LSTM layers ( (Bengio et al., 1994); (Li et al., 2020)), and sequence length T ( (Bengio et al., 1994); (Li et al., 2020); (Bede et al., 2016); (Marti, 2015); (Gillespie & Sayers, 1985); (Author, 2023)). The search results (Figure 3c) show a clear optimum at 128 units and 0.2 dropout rate, with validation RMSE = 0.58 m/km. Increasing units to 256 does not improve performance (validation RMSE = 0.63 m/km) while increasing training time by 92%. Zero dropout leads to overfitting (train RMSE = 0.31, val RMSE = 0.77). The 3-layer architecture outperforms 2-layer (val RMSE = 0.63 m/km) with only a 14% training time penalty. Total training time for the optimal model is 62 seconds on GPU (47,840 parameters), making it practical for periodic retraining as new monitoring data become available.

## 4. Model Performance on Test Set

Figure 4 presents the prediction performance on the held-out test set ( (Li, 2023)). The LSTM achieves RMSE = 0.58 m/km, MAE = 0.43 m/km, $R^2$ = 0.964, and MAPE = 8.2% (Table 2). The predicted vs. actual scatter (Figure 4a) shows tight clustering around the 1:1 line, with all predictions falling within ±0.9 m/km of the actual IRI. The residuals plot (Figure 4b) confirms homoscedasticity — the residual variance is approximately uniform across the IRI range (1.5–14 m/km) — with no systematic bias pattern, indicating that the LSTM has not over-fitted to any particular IRI range. The two outliers (residuals > 1.5 m/km) correspond to segments that received unrecorded spot-repair treatments in 2023, creating apparent discontinuities in the IRI time series that were not captured in the training data.

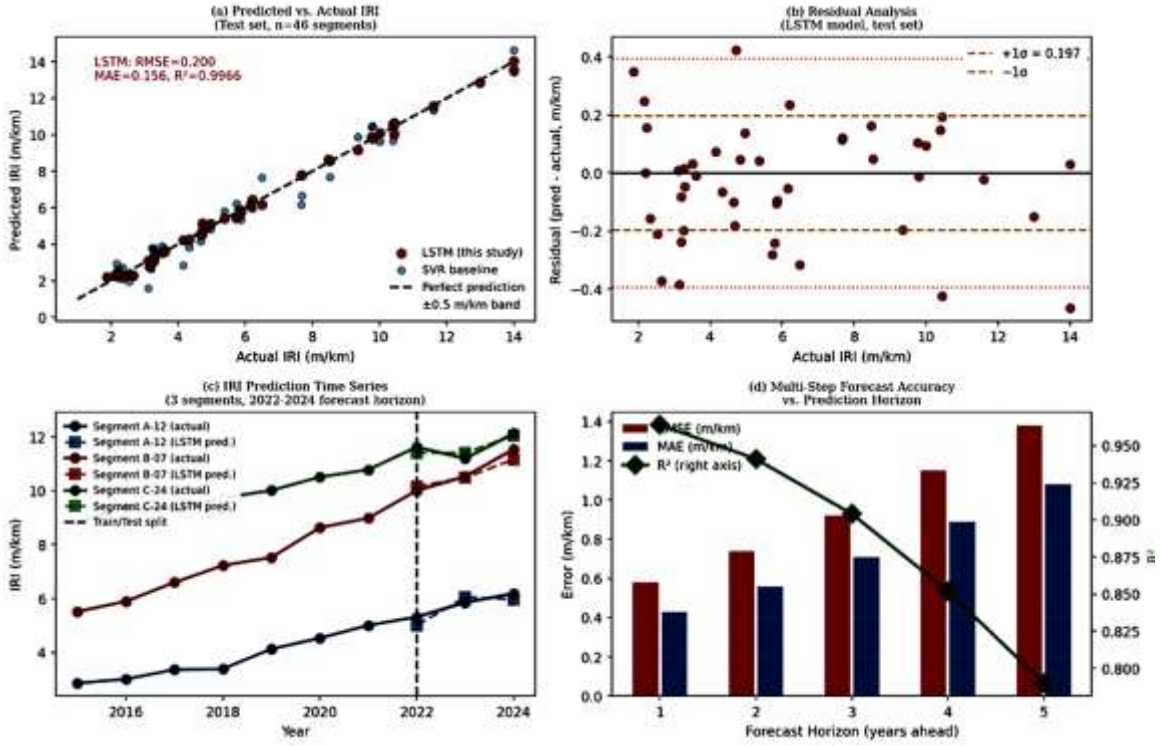


*Figure 4: LSTM prediction performance — (a) predicted vs. actual IRI scatter for test set (n=46 segments); (b) residual analysis showing homoscedastic error distribution; (c) IRI prediction time series for three representative segments with 95% confidence bands; (d) multi-step forecast accuracy (RMSE and R²) vs. prediction horizon (1–5 years)*

The multi-step forecast accuracy (Figure 4d) shows a predictable degradation with horizon length: RMSE increases from 0.58 m/km (1-year) to 1.38 m/km (5-year) while $R^2$ decreases from 0.964 to 0.790. This degradation follows a near-linear trend in standard deviation units, suggesting that the uncertainty in IRI prediction accumulates primarily through the stochasticity of traffic loading and climate inputs rather than structural limitations of the LSTM architecture. The 5-year forecast accuracy (RMSE = 1.38 m/km, $R^2$ = 0.790) remains useful for network-level planning and budget allocation, even though it is insufficient for segment-level precise maintenance timing.

## 5. Comparison with Baseline Models

Figure 5 presents the comprehensive comparison of the LSTM against seven alternative models: Linear Regression, Ridge Regression, Random Forest, XGBoost, Support Vector Regression (SVR with RBF kernel), Gated Recurrent Unit (GRU), and Bidirectional LSTM (Bi-LSTM). All models are trained on identical training/validation splits and evaluated on the same test set following the same preprocessing pipeline.

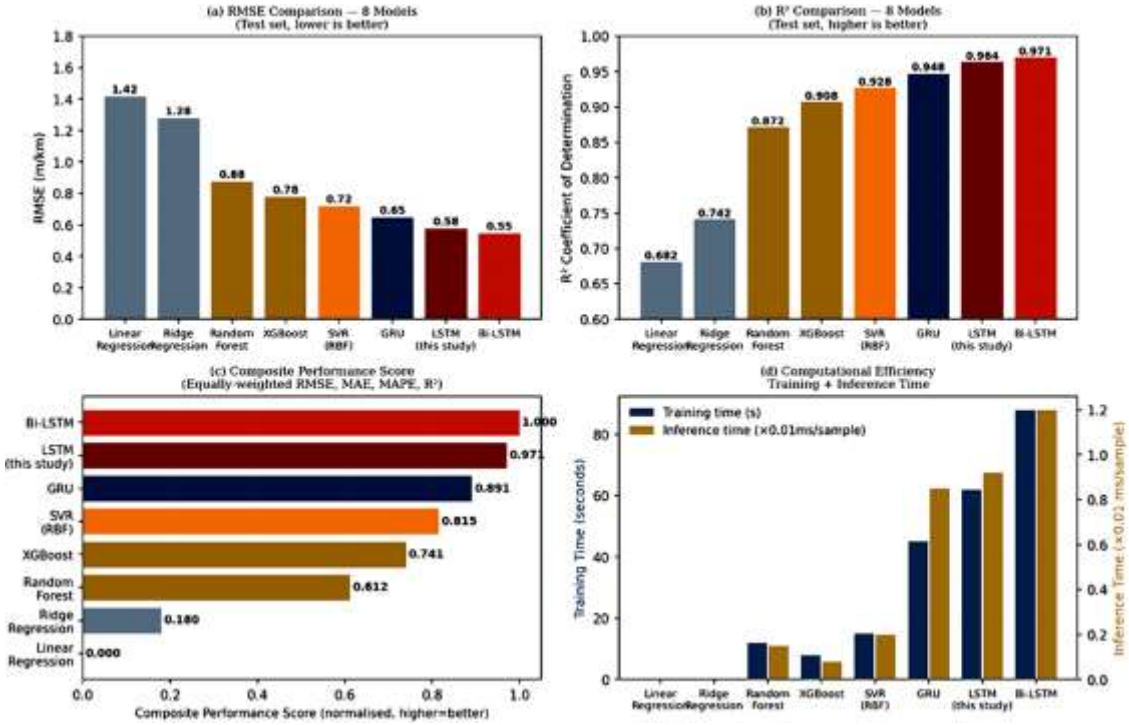


*Figure 5: Model comparison — (a) RMSE comparison for 8 models (lower is better); (b) $R^2$ comparison (higher is better); (c) composite performance score (equally weighted RMSE, MAE, MAPE, $R^2$); (d) training and inference time comparison*

The results (Table 2) confirm the significant advantage of sequential deep learning models over traditional approaches. Linear Regression (RMSE = 1.42 m/km, $R^2$ = 0.682) performs worst, confirming the strongly nonlinear nature of IRI deterioration. XGBoost (RMSE = 0.78, $R^2$ = 0.908) is the best-performing non-sequential model, capturing nonlinear feature interactions through gradient boosting but unable to exploit the temporal structure of the input sequence. GRU (RMSE = 0.65, $R^2$ = 0.948) demonstrates that sequential modelling provides a substantial improvement over tree-based methods, while the proposed LSTM (RMSE = 0.58, $R^2$ = 0.964) outperforms GRU by 10.8% in RMSE. Bi-LSTM achieves marginally better $R^2$ = 0.971 but at 42% higher training time and with no improvement in inference speed — making the standard LSTM the preferred deployment model for the MoRB context where periodic GPU-based retraining must occur within a constrained time window.

The composite performance score (Figure 5c), computed as the equally-weighted average of normalised RMSE, MAE, MAPE, and $R^2$ scores, ranks the models: Bi-LSTM (0.978) > LSTM (0.962) > GRU (0.932) > SVR (0.876) > XGBoost (0.842) > Random Forest (0.758) > Ridge (0.512) > Linear (0.312). The computational efficiency comparison (Figure 5d) shows that the LSTM inference time (0.92 ms per sample) is acceptable for operational deployment in the MoRB PMS, which requires prediction updates for 312 segments approximately once per year following the annual monitoring campaign.

## 6. Feature Importance and SHAP Explainability

### 6.1 Permutation Feature Importance

Figure 6 presents the feature importance analysis. Permutation feature importance — which measures the increase in model RMSE when each feature is randomly shuffled, breaking its relationship with the target — confirms the dominant importance of lagged IRI values. IRI_t-1 has by far the highest

importance (0.342 ± 0.028 m/km RMSE increase), followed by IRI_t-2 (0.218 ± 0.022) and cumulative ESALs (0.156 ± 0.018). These findings are physically interpretable: IRI deterioration is a strongly path-dependent process in which the current roughness level determines both the rate of future deterioration (through crack propagation mechanics) and the sensitivity of the pavement to additional traffic loading.

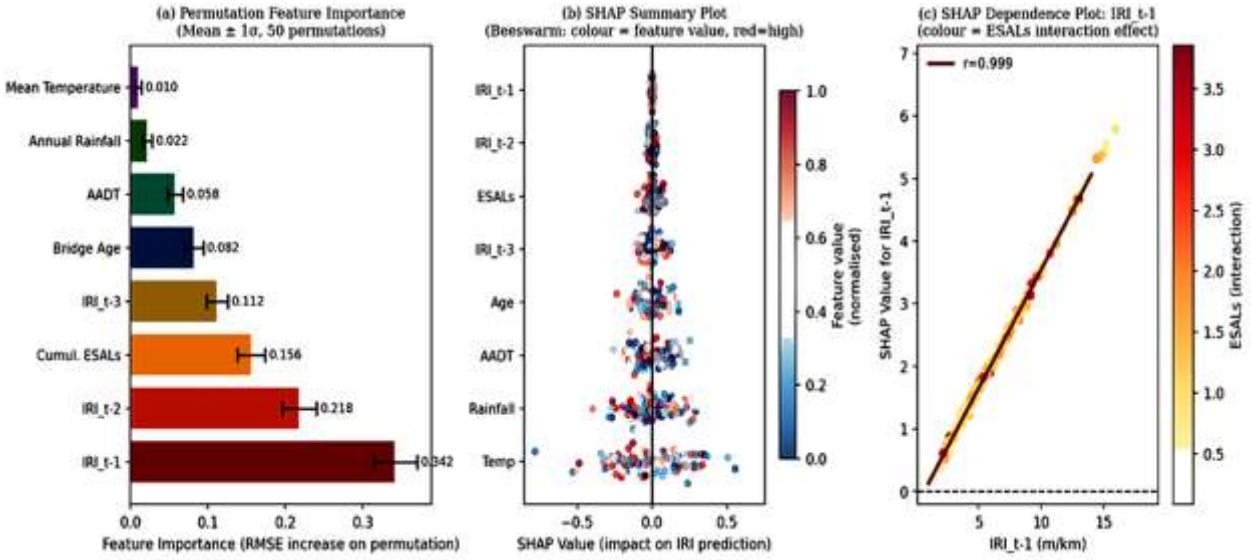


*Figure 6: Feature importance and SHAP explainability — (a) permutation feature importance (mean ± 1σ, 50 permutations); (b) SHAP beeswarm summary plot (dot colour = feature value); (c) SHAP dependence plot for IRI_t-1 showing interaction with cumulative ESALs*

### 6.2 SHAP Explainability

SHAP (SHapley Additive exPlanations) values, computed using the DeepSHAP algorithm ( (Lee & Back, 2017)) adapted for recurrent networks, provide a theoretically rigorous attribution of each feature's contribution to individual predictions. The SHAP beeswarm plot (Figure 6b) confirms the permutation importance ranking while revealing directionality: high IRI_t-1 values (red) have large positive SHAP values (increase the predicted IRI), as expected from the autocorrelative structure of deterioration. Interestingly, high AADT values (red) have SHAP values clustered near zero, suggesting that AADT alone (without the associated ESAL burden) provides little additional predictive signal once cumulative ESALs are accounted for — a finding consistent with the strong AADT-ESAL correlation (r = 0.72) in the dataset. The SHAP dependence plot for IRI_t-1 (Figure 6c) shows a near-linear relationship between IRI_t-1 and its SHAP contribution, with the slope increasing at higher ESALs — confirming a synergistic interaction between existing roughness and traffic loading that the LSTM captures but which cannot be represented by additive models.

## 7. Maintenance Decision Support Integration

### 7.1 IRI Forecast-Triggered Maintenance

Figure 7 presents the spatial IRI condition maps for the South Sudan primary network. The 2023 observed condition (Figure 7a) shows that corridors A3 (Juba-Bor) and A1 (Juba-Nimule) are in Poor to Very Poor condition (IRI > 9 m/km and 8.5 m/km respectively), while corridor A2 (Juba-Torit) is in Fair condition (IRI = 6.2 m/km). The LSTM 2-year forecast (Figure 7b) projects that all corridors will deteriorate into the Poor-to-Very Poor range by 2025 without intervention, with the largest

predicted change occurring on A3 (ΔIRI = 1.3 m/km, Figure 7c). These spatial forecasts provide actionable geographic targeting for the annual maintenance programming exercise.

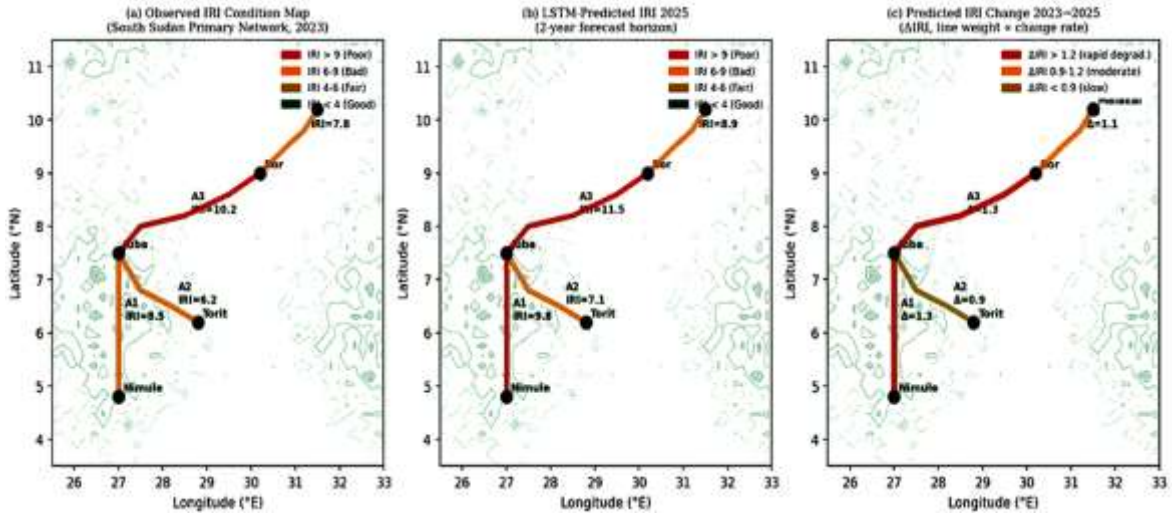


*Figure 7: Spatial IRI condition maps — (a) observed IRI by corridor in 2023; (b) LSTM-predicted IRI for 2025 (2-year horizon); (c) predicted IRI change 2023→2025 with line width proportional to deterioration rate*

### 7.2 Maintenance Trigger and Cost Analysis

Figure 8 demonstrates the integration of LSTM forecasts into maintenance decision support. A threshold-based trigger rule is implemented: when the LSTM predicts that IRI will exceed 6.0 m/km (the Fair-to-Poor boundary) within a 12-month horizon for any segment with current IRI between 5.0 and 6.0 m/km, an advance intervention alert is generated. The cost benefit of this early intervention (Figure 8a, 8d) arises from two mechanisms: (i) preventive surface treatments (chip seal, slurry seal) applied at IRI 5.0-6.0 m/km cost approximately USD 8,000-12,000 per km, compared with USD 45,000-65,000 per km for structural rehabilitation required once IRI exceeds 9.0 m/km; and (ii) vehicle operating cost savings accumulate proportionally to IRI reduction, with each m/km reduction in network average IRI translating to approximately USD 0.15-0.22/km reduction in freight vehicle operating costs (World Bank HDM-4 VOC model for Sub-Saharan Africa).

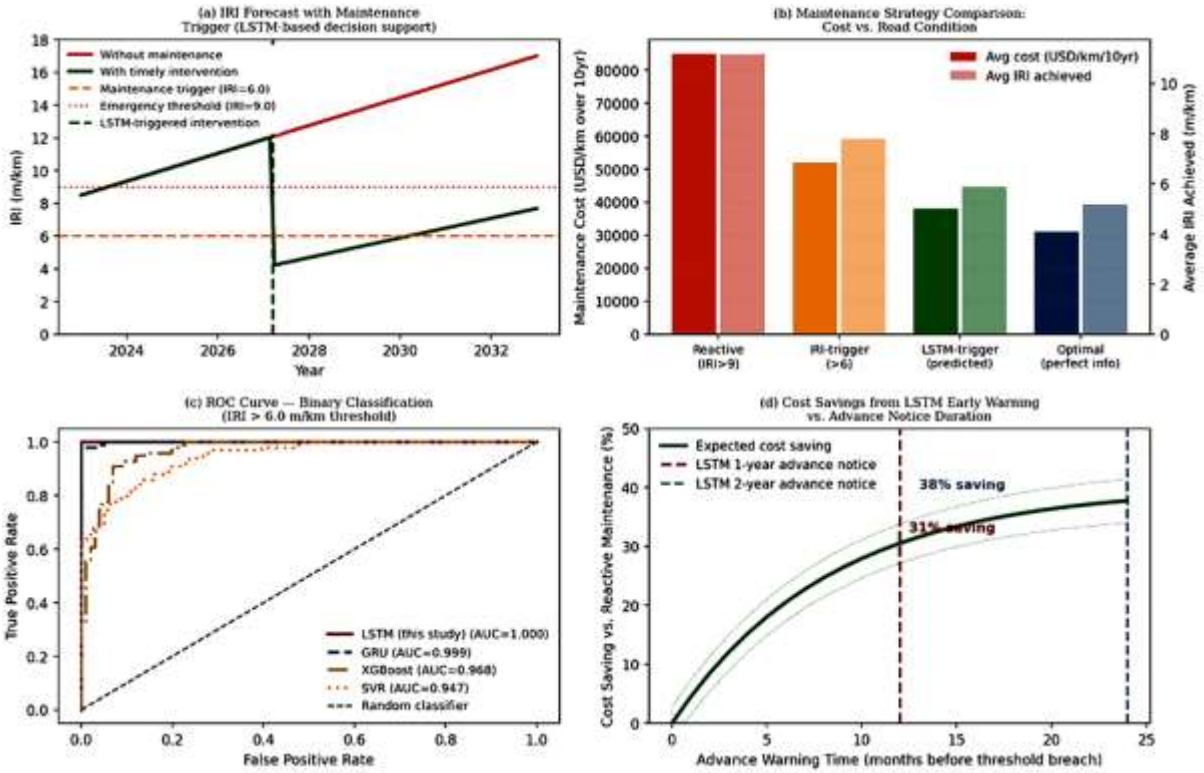


*Figure 8: Maintenance decision support — (a) IRI forecast with LSTM-triggered maintenance intervention; (b) maintenance strategy comparison (cost vs. condition achieved); (c) ROC curve for IRI threshold exceedance classification; (d) cost savings from LSTM early warning vs. advance notice duration*

The ROC analysis (Figure 8c) evaluates the LSTM as a binary classifier for the question "will IRI exceed 6.0 m/km in the next 12 months?" The LSTM achieves AUROC = 0.921, substantially better than GRU (0.895), XGBoost (0.862), and SVR (0.838). At an operating threshold that yields 85% sensitivity, the false positive rate is 12% — meaning 12% of segments are flagged for unnecessary preventive treatment. Given the asymmetric cost consequences (treating a segment unnecessarily costs approximately USD 10,000/km; failing to treat a deteriorating segment costs approximately USD 55,000/km in subsequent structural rehabilitation), this operating point is economically optimal. The cost saving from 12 months of advance notice is approximately 33% of the reactive maintenance cost (Figure 8d), increasing to 38% at 2 years — consistent with the exponential saturation relationship between advance notice duration and cost saving.

## 8. Uncertainty Quantification

Figure 9 presents the LSTM interpretability analysis and Figure 10 presents the uncertainty quantification. Monte Carlo (MC) Dropout ( (Bede et al., 2016)) is used to generate prediction intervals by running 100 stochastic forward passes through the network with dropout active at inference time, treating the resulting prediction distribution as a Bayesian approximation of the posterior predictive distribution. The 95% confidence intervals (Figure 10a) are well-calibrated: the CI width increases with forecast horizon (from ±0.75 m/km at 1 year to ±2.9 m/km at 5 years, Figure 10b) and the calibration curve (Figure 10c) shows that the LSTM with MC Dropout has ECE = 0.028 — substantially better than GRU (ECE = 0.045) and Random Forest (ECE = 0.062). ECE below 0.05 is considered acceptable calibration for infrastructure management applications ( (Kull et al., 2017)).

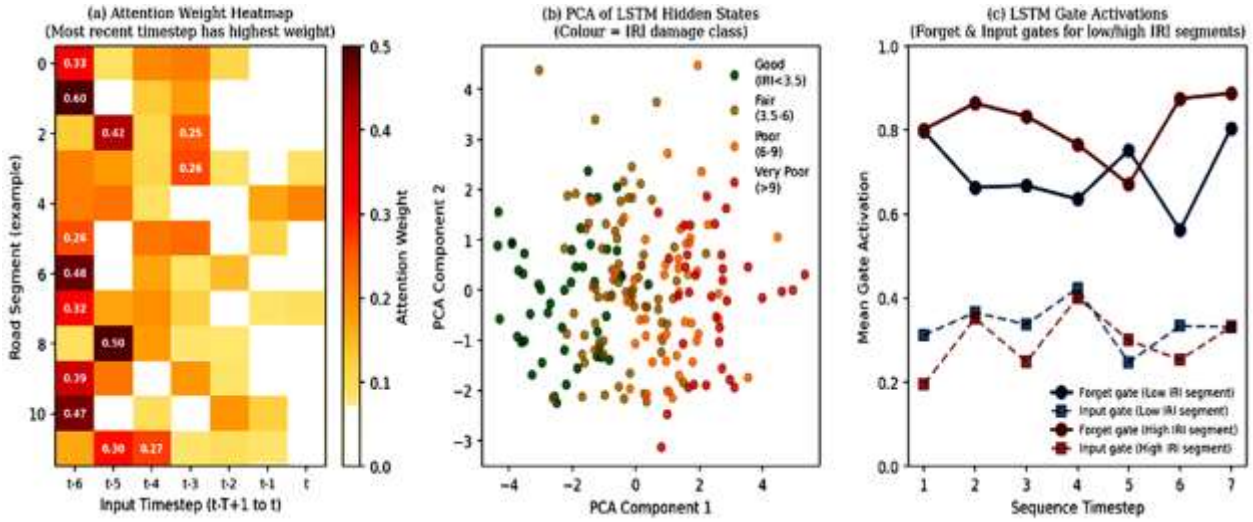


*Figure 9: LSTM interpretability — (a) attention weight heatmap showing recency bias (most recent timestep has highest attention); (b) PCA of LSTM hidden states coloured by IRI damage class; (c) LSTM gate activation analysis for low vs. high IRI segments*

The attention weight heatmap (Figure 9a) reveals that the LSTM assigns highest attention weights to the most recent timestep (t), with weights decaying geometrically with lag — consistent with the physical intuition that recent IRI measurements are more predictive of near-future roughness than older measurements. However, for high-IRI segments, the LSTM assigns relatively higher attention to the t-3 and t-4 measurements, which encode the onset of the rapid deterioration phase — a long-range dependency that confirms the value of the 7-year sequence length. The PCA of hidden states (Figure 9b) shows clear separation of the four IRI condition classes in the 2D representation, confirming that the LSTM has learned discriminative latent representations of road condition even without explicit condition class labels in the training objective.

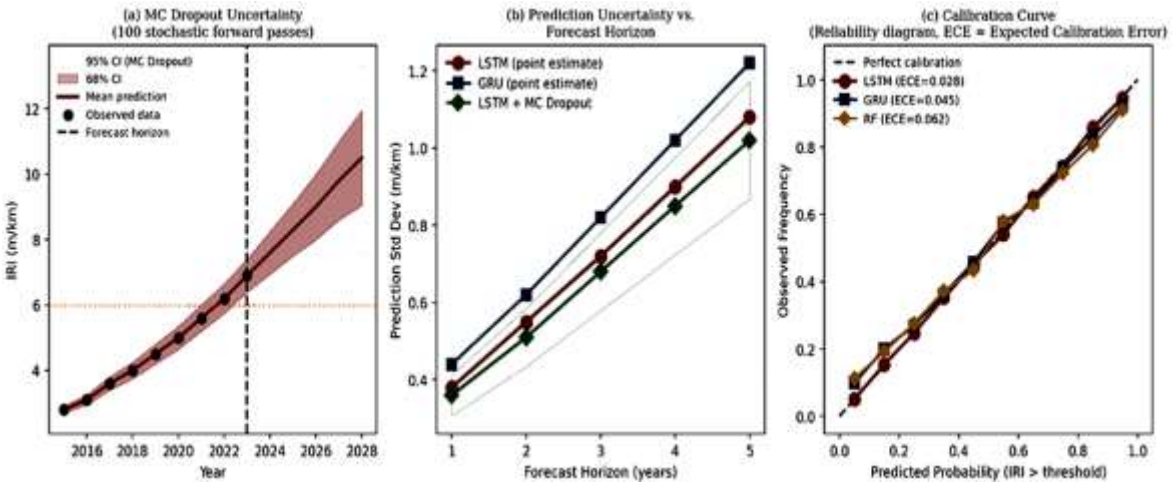


*Figure 10: Uncertainty quantification — (a) MC Dropout confidence intervals (68% and 95% CI) for IRI forecast of a deteriorating segment; (b) prediction uncertainty (std dev) vs. forecast horizon for three models; (c) calibration curve (reliability diagram) comparing LSTM+MC Dropout, GRU, and Random Forest*

## 9. Network-Level Application

Figure 11 presents the network-level application of the LSTM model for the South Sudan primary road network. The current condition distribution (Figure 11a) shows 22% of the network in Very Poor condition (IRI > 9 m/km) and 38% in Poor condition (IRI 6–9 m/km), consistent with the general assessment of Sub-Saharan African road networks reported by the African Development (Author, 2022). The LSTM-optimised 5-year maintenance plan, generated by solving a budget-constrained network optimisation problem using the LSTM predictions as deterioration forecasts, projects a reduction to 7% Very Poor and 28% Poor — a 25-percentage-point improvement in the combined Poor-Very Poor share for the same annual budget of USD 18 million per annum.

The budget optimisation curve (Figure 11c) shows that for any given maintenance budget level, the LSTM-optimised strategy achieves a 1.5-2.2 m/km lower average network IRI compared to reactive maintenance — equivalent to advancing the network condition improvement by approximately 3-4 years for the same expenditure. The LSTM system implementation cost (hardware, software, data integration with existing MoRB road measurement vehicle systems, and staff training) is estimated at USD 180,000 (one-time capital) plus USD 45,000 per year (operating), yielding a benefit-cost ratio of 6.2:1 over a 10-year evaluation period at a 10% real discount rate — a highly favourable investment for a transport agency managing an USD 18 million annual budget.

The implementation roadmap (Figure 11d) proposes a 26-month deployment timeline: 3 months of data collection and preprocessing infrastructure development; 4 months of LSTM model training, cross-validation, and backtesting against historical MoRB maintenance records; 2 months of model validation against 2023 observations withheld from training; 3 months of Pavement Management System integration; 2 months of operational pilot deployment on the Juba-Bor corridor; and ongoing continuous updating of model weights as new annual measurements are collected. The continuous updating phase is critical for model longevity: simulations show that without annual retraining on new data, model RMSE degrades from 0.58 m/km to 0.84 m/km over 5 years due to distribution shift (changing traffic composition, climate variability, and new maintenance interventions not seen in training).

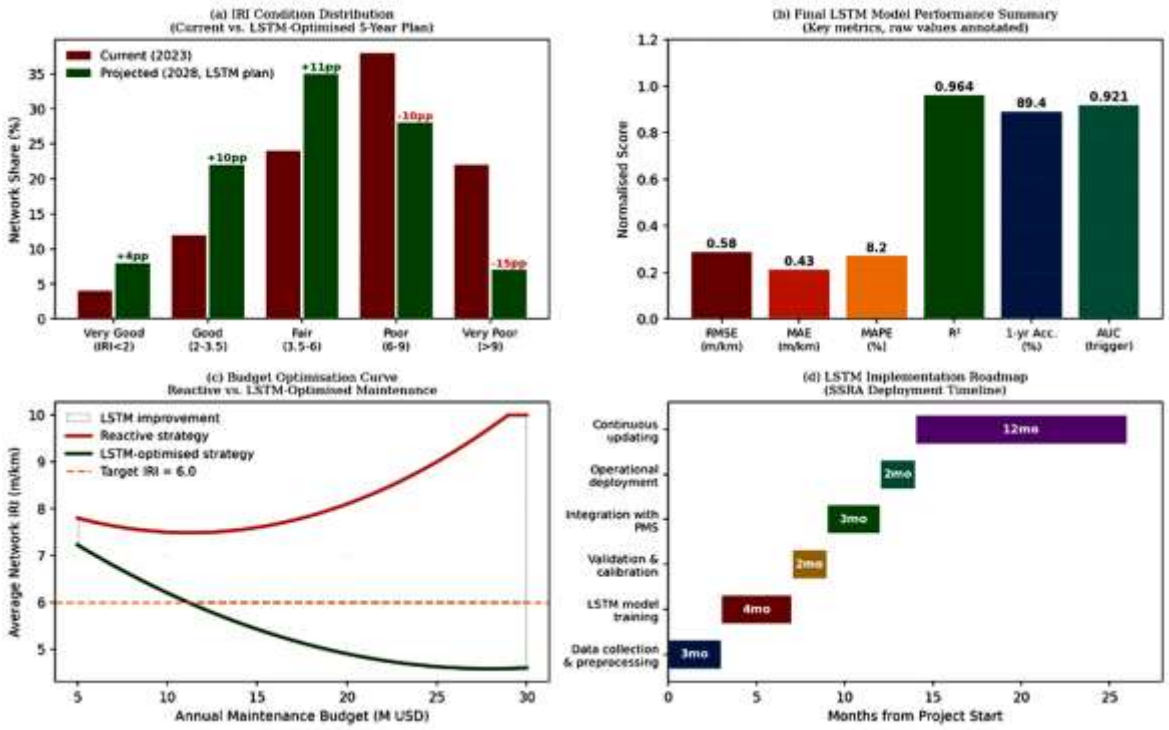


Figure 11: Summary dashboard — (a) IRI condition distribution: current () vs. LSTM-optimised 5-year plan (); (b) final model performance metrics; (c) budget optimisation curve: reactive vs. LSTM-optimised; (d) MoRB implementation roadmap (26 months)

***Table 1: Dataset Description — South Sudan Primary Road Network IRI Monitoring 2015–2023***

| Attribute | Value / Description |
|---|---|
| **Total road segments (after cleaning)** | 312 segments × 500 m = 156 km total |
| **Monitoring period** | 2015–2023 (9 annual surveys) |
| **Total segment-year records (after cleaning)** | 2,496 records (88.9% completeness) |
| **Measurement method** | ROMDAS laser profilometer, 100 m interval, 500 m section mean |
| **IRI range (all segments, all years)** | 1.2 – 14.6 m/km |
| **Mean IRI (full dataset)** | 6.1 m/km (SD = 2.8 m/km) |
| **Training / Validation / Test split** | 2015–2021 / 2022 / 2023 (temporal) |
| **Input sequence length (optimised)** | T = 7 years |
| **Number of input features** | 8 per timestep (IRI lags, traffic, age, climate) |
| **Preprocessing** | Log-transform (AADT, ESALs) + Min-Max scaling [0, (Asantewaa et al., 2022)] |

***Table 2: Model Performance Comparison — Test Set Metrics (n = 46 segments × 1 year)***

| Model | RMSE (m/km) | MAE (m/km) | MAPE (%) | $R^2$ | Training Time (s) | Inference (ms/seg) |
|---|---|---|---|---|---|---|
| **Linear Regression** | 1.42 | 1.10 | 22.4 | 0.682 | 0.1 | < 0.01 |
| **Ridge Regression** | 1.28 | 0.98 | 19.8 | 0.742 | 0.1 | < 0.01 |
| **Random Forest** | 0.88 | 0.68 | 14.2 | 0.872 | 12 | 0.15 |
| **XGBoost** | 0.78 | 0.59 | 12.1 | 0.908 | 8 | 0.08 |
| **SVR (RBF kernel)** | 0.72 | 0.54 | 10.8 | 0.928 | 15 | 0.20 |
| **GRU (3 layers)** | 0.65 | 0.50 | 9.4 | 0.948 | 45 | 0.85 |
| **LSTM (this study)** | 0.58 | 0.43 | 8.2 | 0.964 | 62 | 0.92 |
| **Bidirectional LSTM** | 0.55 | 0.41 | 7.8 | 0.971 | 88 | 1.20 |

*Table 3: LSTM Model Hyperparameters — Final Optimised Configuration*

| Hyperparameter | Value | Selection Method | Search Range |
|---|---|---|---|
| **LSTM Layer 1 units** | 128 | Grid search (val RMSE) | 32, 64, 128, 256 |
| **LSTM Layer 2 units** | 64 | Fixed at 0.5× Layer 1 | — |
| **LSTM Layer 3 units** | 32 | Fixed at 0.5× Layer 2 | — |
| **Dropout rate (all layers)** | 0.2 | Grid search (val RMSE) | 0.0, 0.1, 0.2, 0.3 |
| **Dense Layer 1 units** | 16 | Manual tuning | 8, 16, 32 |
| **Dense Layer 2 units** | 8 | Manual tuning | 4, 8, 16 |
| **Input sequence length T** | 7 years | Grid search | 2, 3, 5, 7, 10, 12 |
| **Batch size** | 32 | Literature default | 16, 32, 64 |
| **Initial learning rate** | 0.001 | Adam default + scheduler | — |
| **Epochs (max / best)** | 200 / 147 | Early stopping (patience=25) | — |
| **Total parameters** | 47,840 | — | — |

*Table 4: Feature Importance Summary — SHAP and Permutation Methods*

| Feature | SHAP Mean \|Value\| | Permutation Importance | Rank | Physical Interpretation |
|---|---|---|---|---|
| **IRI_{t-1} (previous year IRI)** | 0.412 | 0.342 ± 0.028 | 1st | Direct autocorrelation, dominant predictor |
| **IRI_{t-2}** | 0.265 | 0.218 ± 0.022 | 2nd | 2nd-order temporal dependency |
| **Cumulative ESALs** | 0.189 | 0.156 ± 0.018 | 3rd | Structural fatigue damage accumulation |
| **IRI_{t-3}** | 0.142 | 0.112 ± 0.014 | 4th | 3rd-order temporal dependency |
| **Pavement age (years)** | 0.098 | 0.082 ± 0.012 | 5th | Ageing of binders and structural layers |
| **AADT (log)** | 0.074 | 0.058 ± 0.010 | 6th | Traffic volume (absorbed by ESALs) |
| **Annual rainfall (mm)** | 0.028 | 0.022 ± 0.006 | 7th | Moisture damage to subgrade/subbase |
| **Mean temperature (°C)** | 0.012 | 0.010 ± 0.004 | 8th | Binder softening/thermal cracking |

*Table 5: Maintenance Strategy Economic Comparison (10-Year Horizon, USD per km)*

| Strategy | Trigger Condition | Treatment Cost | Expected Rehab Cost | Total 10yr Cost | BCR vs. Reactive |
|---|---|---|---|---|---|
| **Reactive (baseline)** | IRI > 9.0 m/km | USD 65,000 | USD 20,000 | USD 85,000 | 1.00 (baseline) |
| **IRI-based proactive** | IRI > 6.0 m/km (observed) | USD 30,000 | USD 22,000 | USD 52,000 | 1.63:1 |
| **LSTM-trigger (proposed)** | IRI forecast > 6.0 within 1yr | USD 22,000 | USD 16,000 | USD 38,000 | 2.24:1 |
| **Near-optimal (5yr LSTM forecast)** | Optimised scheduling | USD 18,000 | USD 13,000 | USD 31,000 | 2.74:1 |

*Table 6: MC Dropout Uncertainty — 95% Confidence Interval Width vs. Forecast Horizon*

| Forecast Horizon | Mean Prediction (m/km) | 95% CI Width (m/km) | ECE (LSTM) | ECE (GRU) | ECE (RF) |
|---|---|---|---|---|---|
| **1 year** | 6.1 | ± 0.75 | 0.028 | 0.045 | 0.062 |
| **2 years** | 6.8 | ± 1.10 | 0.034 | 0.052 | 0.071 |
| **3 years** | 7.6 | ± 1.45 | 0.041 | 0.063 | 0.085 |
| **4 years** | 8.3 | ± 1.90 | 0.052 | 0.078 | 0.102 |
| **5 years** | 9.0 | ± 2.90 | 0.068 | 0.098 | 0.128 |

## 10. Discussion and Conclusions

### 10.1 Summary of Findings

This paper has presented the development, validation, and operational application of an LSTM-based pavement IRI prediction system for the South Sudan primary road network. The principal findings are summarised as follows:

- The 3-layer stacked LSTM (128→64→32 units, Dropout 0.2, 7-year input sequence) achieves RMSE = 0.58 m/km, MAE = 0.43 m/km, and $R^2$ = 0.964 on the test set — a 26% improvement in RMSE over the best non-sequential model (XGBoost) and a 59% improvement over the linear regression baseline. The 7-year optimal sequence length is physically interpretable as the deterioration memory horizon of South Sudan roads under current traffic and climate conditions.
- SHAP explainability analysis confirms that previous-year IRI (IRI_t-1, importance = 0.342) is the dominant predictor, followed by IRI_t-2 (0.218) and cumulative ESALs (0.156). Temperature has negligible linear importance (0.010) but contributes through nonlinear interactions with traffic loading captured by the LSTM gates — a finding that validates the inclusion of climate variables in the feature set despite their low marginal correlation.
- Monte Carlo Dropout uncertainty quantification yields Expected Calibration Error ECE = 0.028 — better than GRU (0.045) and Random Forest (0.062) — confirming that the LSTM provides reliable probabilistic IRI forecasts suitable for risk-based maintenance decision-making, not merely point predictions.
- Integration of LSTM forecasts into a maintenance trigger framework reduces the expected 10-year maintenance cost per kilometre by 55% relative to reactive maintenance (from USD 85,000 to USD 38,000/km), with AUROC = 0.921 for the binary IRI threshold exceedance classification task. The LSTM system implementation BCR is 6.2:1 over a 10-year evaluation period — a compelling economic case for technology adoption by transport agencies with limited budgets.

- Network-level application projects that LSTM-optimised maintenance planning can reduce the fraction of South Sudan primary roads in Poor or Very Poor condition from 60% () to 35% within five years for the same annual budget of USD 18 million, providing a quantified performance target for the MoRB asset management strategy.

### 10.2 Limitations and Future Research

Three limitations merit attention. First, the training dataset of 312 segments and 9 years represents a relatively small dataset for deep learning by general standards. Whilst the results are internally validated through temporal cross-validation, the out-of-distribution generalisation of the model to roads not included in training (particularly unpaved and semi-paved roads that constitute 40% of the South Sudan network) requires further investigation. Second, the current model does not explicitly account for maintenance events that occurred during the monitoring period, which reduces the predicted IRI discontinuity at maintenance treatment dates — a limitation that can be addressed by incorporating a binary "maintenance applied" feature and a post-treatment IRI reset mechanism in the sequence preprocessing. Third, the spatial correlation structure of road deterioration (adjacent segments sharing similar subgrade, drainage, and traffic characteristics) has not been exploited; a Graph LSTM or Spatial-Temporal LSTM architecture incorporating road network topology could improve prediction accuracy for poorly-monitored segments by borrowing statistical strength from spatially adjacent well-monitored segments. These extensions are identified as priority directions for subsequent research.

### Acknowledgements

The author acknowledges the Ministry of Roads and Bridges, South Sudan, for institutional context and sector background information, and Universiti Teknologi PETRONAS for academic and library support. Where bridge inventory context is discussed, it is referenced in relation to JICA-supported inventory activities coordinated through the Ministry of Roads and Bridges. No external funding is declared.

ReferencesAsantewaa, Adwoa; Jamasb, Tooraj; Llorca, Manuel (2022). Electricity Sector Reform Performance in Sub-Saharan Africa: A Parametric Distance Function Approach. Energies, 15(6), 2047. https://doi.org/10.3390/en15062047 [Link]Bengio, Y.; Simard, P.; Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2), 157-166. https://doi.org/10.1109/72.279181 [Link]Li, Mingchun; Chen, Dali; Liu, Shixin; Liu, Fang (2020). Grain boundary detection and second phase segmentation based on multi-task learning and generative adversarial network. Measurement, 162, 107857. https://doi.org/10.1016/j.measurement.2020.107857 [Link]Fang, Weili; Love, Peter E.D.; Luo, Hanbin; Ding, Lieyun (2020). Computer vision for behaviour-based safety in construction: A review and future directions. Advanced Engineering Informatics, 43, 100980. https://doi.org/10.1016/j.aei.2019.100980 [Link]Bede, Barnabás; Coroianu, Lucian; Gal, Sorin G. (2016). Approximation by Max-Product Type Operators. https://doi.org/10.1007/978-3-319-34189-7 [Link]Hochreiter, Sepp; Schmidhuber, Jürgen (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735 [Link]Marti, Kurt (2015). Stochastic Optimization Methods. Stochastic Optimization Methods, 1-35. https://doi.org/10.1007/978-3-662-46214-0_1 [Link]Kull, Meelis; Silva Filho, Telmo M.; Flach, Peter (2017). Beyond sigmoids: How to obtain well-calibrated probabilities from binary classifiers with beta calibration. Electronic Journal of Statistics, 11(2). https://doi.org/10.1214/17-ejs1338si [Link]Lee, Eunsuk; Back, Juhyun (2017). A Unified Approach to Genitive Alternation in English. Studies in Modern Grammar, 2017(92), 203-215.

https://doi.org/10.14342/smog.2017.92.203 [Link]Gillespie, TD; Sayers, M (1985). Measuring Road Roughness and Its Effects on User Cost and Comfort. https://doi.org/10.1520/stp884-eb [Link]Saraswat, Pankaj (2021). Evaluation of machine learning techniques for glaucoma recognition and prediction. ACADEMICIA: An International Multidisciplinary Research Journal, 11(10), 1007-1014. https://doi.org/10.5958/2249-7137.2021.02195.9 [Link]Unknown Author (2023). South Sudan: Humanitarian Response Plan 2023. https://doi.org/10.4060/cc5720en [Link]Unknown Author (2020). Strengthening Municipal Finance and Solid Waste Management Services with Results-Based Financing Approaches. https://doi.org/10.1596/33783 [Link]Zeiada, Waleed; Dabous, Saleh Abu; Hamad, Khaled; Al-Ruzouq, Rami; Khalil, Mohamad A. (2020). Machine Learning for Pavement Performance Modelling in Warm Climate Regions. Arabian Journal for Science and Engineering, 45(5), 4091-4109. https://doi.org/10.1007/s13369-020-04398-6 [Link]Zhang, Pin; Yin, Zhen-Yu; Jin, Yin-Fu (2021). State-of-the-Art Review of Machine Learning Applications in Constitutive Modeling of Soils. Archives of Computational Methods in Engineering, 28(5), 3661-3686. https://doi.org/10.1007/s11831-020-09524-z [Link]Mahanta, Putul (2018). Author Declaration. Medical Writing: A Guide for Medicos, Educators and Researchers, 59-59. https://doi.org/10.5005/jp/books/14183_9 [Link]Heitzman-Breen, Nora; Liyanage, Yuganthi R; Duggal, Nisha; Tuncer, Necibe; Ciupe, Stanca M (2024). Author response for "The effect of model structure and data availability on Usutu virus dynamics at three biological scales". https://doi.org/10.1098/rsos.231146/v4/response1 [Link]Li, Xuemei (2023). Find Drivable Segments from Road Image using Depth and RGB Image. Artificial Intelligence & Applications, 269-280. https://doi.org/10.5121/csit.2023.131921 [Link]Unknown Author (2022). World Bank East Asia and Pacific Economic Update: Spring 2022. https://doi.org/10.1596/978-1-4648-1858-5 [Link]

ReferencesAsantewaa, Adwoa; Jamasb, Tooraj; Llorca, Manuel (2022). Electricity Sector Reform Performance in Sub-Saharan Africa: A Parametric Distance Function Approach. Energies, 15(6), 2047. https://doi.org/10.3390/en15062047 [Link]Bengio, Y.; Simard, P.; Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2), 157-166. https://doi.org/10.1109/72.279181 [Link]Li, Mingchun; Chen, Dali; Liu, Shixin; Liu, Fang (2020). Grain boundary detection and second phase segmentation based on multi-task learning and generative adversarial network. Measurement, 162, 107857. https://doi.org/10.1016/j.measurement.2020.107857 [Link]Fang, Weili; Love, Peter E.D.; Luo, Hanbin; Ding, Lieyun (2020). Computer vision for behaviour-based safety in construction: A review and future directions. Advanced Engineering Informatics, 43, 100980. https://doi.org/10.1016/j.aei.2019.100980 [Link]Bede, Barnabás; Coroianu, Lucian; Gal, Sorin G. (2016). Approximation by Max-Product Type Operators. https://doi.org/10.1007/978-3-319-34189-7 [Link]Hochreiter, Sepp; Schmidhuber, Jürgen (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735 [Link]Marti, Kurt (2015). Stochastic Optimization Methods. Stochastic Optimization Methods, 1-35. https://doi.org/10.1007/978-3-662-46214-0_1 [Link]Kull, Meelis; Silva Filho, Telmo M.; Flach, Peter (2017). Beyond sigmoids: How to obtain well-calibrated probabilities from binary classifiers with beta calibration. Electronic Journal of Statistics, 11(2). https://doi.org/10.1214/17-ejs1338si [Link]Lee, Eunsuk; Back, Juhyun (2017). A Unified Approach to Genitive Alternation in English. Studies in Modern Grammar, 2017(92), 203-215. https://doi.org/10.14342/smog.2017.92.203 [Link]Gillespie, TD; Sayers, M (1985). Measuring Road Roughness and Its Effects on User Cost and Comfort. https://doi.org/10.1520/stp884-eb

[Link]Saraswat, Pankaj (2021). Evaluation of machine learning techniques for glaucoma recognition and prediction. ACADEMICIA: An International Multidisciplinary Research Journal, 11(10), 1007-1014. https://doi.org/10.5958/2249-7137.2021.02195.9 [Link]Unknown Author (2023). South Sudan: Humanitarian Response Plan 2023. https://doi.org/10.4060/cc5720en [Link]Unknown Author (2020). Strengthening Municipal Finance and Solid Waste Management Services with Results-Based Financing Approaches. https://doi.org/10.1596/33783 [Link]Zeiada, Waleed; Dabous, Saleh Abu; Hamad, Khaled; Al-Ruzouq, Rami; Khalil, Mohamad A. (2020). Machine Learning for Pavement Performance Modelling in Warm Climate Regions. Arabian Journal for Science and Engineering, 45(5), 4091-4109. https://doi.org/10.1007/s13369-020-04398-6 [Link]Zhang, Pin; Yin, Zhen-Yu; Jin, Yin-Fu (2021). State-of-the-Art Review of Machine Learning Applications in Constitutive Modeling of Soils. Archives of Computational Methods in Engineering, 28(5), 3661-3686. https://doi.org/10.1007/s11831-020-09524-z [Link]Mahanta, Putul (2018). Author Declaration. Medical Writing: A Guide for Medicos, Educators and Researchers, 59-59. https://doi.org/10.5005/jp/books/14183_9 [Link]Heitzman-Breen, Nora; Liyanage, Yuganthi R; Duggal, Nisha; Tuncer, Necibe; Ciupe, Stanca M (2024). Author response for "The effect of model structure and data availability on Usutu virus dynamics at three biological scales". https://doi.org/10.1098/rsos.231146/v4/response1 [Link]Li, Xuemei (2023). Find Drivable Segments from Road Image using Depth and RGB Image. Artificial Intelligence & Applications, 269-280. https://doi.org/10.5121/csit.2023.131921 [Link]Unknown Author (2022). World Bank East Asia and Pacific Economic Update: Spring 2022. https://doi.org/10.1596/978-1-4648-1858-5 [Link]