

# Machine Learning-Based Road Condition Assessment from Satellite Imagery in South Sudan

**Aduot Madit Anhiem**

Research Affiliation: UNICAF / Liverpool John Moores University, Liverpool, UK; UniAthena /  
Guglielmo Marconi University, Rome, Italy

Email: [aduot.madit2022@gmail.com](mailto:aduot.madit2022@gmail.com) | [righker@gmail.com](mailto:righker@gmail.com)

## ABSTRACT

Systematic road condition assessment is a prerequisite for rational maintenance programming and rehabilitation investment decisions, yet conventional field survey methods are prohibitively expensive, logistically constrained, and inaccessible in large areas of South Sudan due to insecurity and seasonal flooding. This paper presents a machine learning (ML) framework for automated road condition assessment using freely available multi-spectral satellite imagery, applied to the classified road network of South Sudan. Six ML models are evaluated — Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Convolutional Neural Network (ResNet-50 architecture), a hybrid Convolutional Neural Network–Long Short-Term Memory (CNN+LSTM) model for temporal feature fusion, and Logistic Regression as a baseline. Input features comprise 24 spectral and textural variables derived from Sentinel-2 Level-2A imagery (10–20 m resolution), Planet NICFI high-resolution basemaps (4.77 m resolution), and derived indices including the Normalised Difference Built-up Index (NDBI), Bare Soil Index (BSI), Modified Normalised Difference Water Index (MNDWI), and Gray-Level Co-occurrence Matrix (GLCM) texture features. Ground truth Road Condition Index (RCI) labels were derived from 1,660 road segments surveyed by the Ministry of Roads and Bridges using standard visual and measurement protocols during February–April 2023. The CNN+LSTM model achieves the highest performance with an Overall Accuracy of 93.5%, Cohen's Kappa of 0.899, and macro-averaged F1 score of 0.922, outperforming XGBoost (89.2%, 0.843, 0.881) and Random Forest (87.4%, 0.821, 0.863). A predicted RCI map for the full classified network (approximately 8,400 km) is generated, revealing that 64% of the network falls in the Poor or Very Poor condition category. Temporal analysis of predicted RCI values from 2019 to 2024 using Sentinel-2 time-series quantifies deterioration rates for three priority corridors. The proposed framework reduces road condition assessment costs by an estimated 87% compared to conventional field surveys and enables annual monitoring cycles feasible within typical government budget envelopes.

**Keywords:** *machine learning; remote sensing; road condition index; Sentinel-2; CNN; LSTM; South Sudan; pavement assessment; satellite imagery; XGBoost*

## 1. INTRODUCTION

---

The reliable assessment of road surface condition is fundamental to infrastructure asset management. Without accurate, up-to-date condition data, maintenance resources are inevitably misallocated — either being applied reactively after catastrophic failure, or distributed uniformly across the network regardless of differential need. In well-resourced road authorities, systematic condition assessment is conducted annually or bi-annually using standardised protocols combining visual survey, automated road analyser vehicles (ARAVs), and roughness measurement with laser profilometers. In South Sudan, however, these conventional approaches are effectively impractical for most of the classified road network: annual field survey coverage by MoRB has historically reached only 12–18% of the network due to insecurity, inaccessibility during the wet season, and a chronic shortage of survey equipment and trained personnel (MoRB, 2022).

The consequence is a national road asset management system operating largely in informational darkness. Rehabilitation investment decisions are made on the basis of stale condition data, political lobbying, and donor preferences rather than objective, network-wide condition assessments. The South Sudan Infrastructure Development Authority (SIDA) has estimated that this informational deficit leads to suboptimal maintenance budget allocation causing approximately 25–30% excess lifecycle costs across the road network (SIDA, 2021). The development of a reliable, low-cost, and scalable road condition assessment methodology is therefore both a technical priority and a governance imperative.

Earth observation (EO) satellite imagery offers a potential solution. The increasing availability of free or low-cost high-resolution multispectral imagery — notably ESA's Sentinel-2 constellation (10–20 m spatial resolution, 5-day revisit), Planet's NICFI basemaps (4.77 m resolution, monthly), and Google Earth Engine's cloud computing platform — makes it technologically feasible to derive road surface condition indicators across entire national networks from desktop environments without fieldwork. Machine learning classification methods can then translate spectral and textural image features into road condition categories or continuous RCI estimates, leveraging the spatial and temporal information content of multi-date image stacks in ways that classical remote sensing indices cannot.

Several recent studies have demonstrated the viability of ML-based road condition assessment from satellite and aerial imagery in data-rich settings, including work by Maas and Rottensteiner (2016) using airborne LiDAR and multispectral data in Germany, Arya et al. (2021) applying deep learning to street-level imagery in India, and Owusu et al. (2021) using random forest classification on Sentinel-2 imagery in Ghana. However, applications specific to conflict-affected, data-scarce Sub-Saharan African

environments — where ground truth data are sparse and the spectral complexity of degraded unpaved and gravel roads in tropical settings poses additional challenges — have not been reported in the literature.

This paper makes the following contributions: (i) it develops and evaluates six ML models for four-class road condition classification using a 24-feature input derived from multispectral satellite imagery calibrated for South Sudan's road network; (ii) it applies a novel CNN+LSTM architecture that fuses spatial convolutional features with temporal sequence learning across six annual Sentinel-2 image composites (2019–2024) to exploit multi-temporal deterioration signals; (iii) it generates the first ML-predicted road condition map for the entire South Sudan classified network; and (iv) it quantifies temporal RCI deterioration trajectories for three strategic corridors, providing data directly applicable to maintenance programming.

The paper is structured as follows. Section 2 reviews relevant literature on ML-based road condition assessment from remote sensing. Section 3 describes the study area, data sources, and ground truth collection. Section 4 details the feature engineering methodology. Section 5 presents the ML model architectures and training procedures. Section 6 reports classification performance results and model comparison. Section 7 presents the predicted RCI map and temporal analysis. Section 8 discusses limitations and future work. Section 9 concludes.

## **2. LITERATURE REVIEW**

---

### **2.1 Remote Sensing for Road Condition Assessment**

Road surface condition assessment from remote sensing has been pursued through several methodological approaches. Early work relied on thermal and multispectral aerial photography to detect surface distress features such as cracking and rutting (Brimley et al., 1996), but the spatial resolution limitations of early satellite platforms (30 m for Landsat TM) restricted applicability to detection of major surface failures rather than subtle condition gradations. With the advent of very-high-resolution commercial imagery (QuickBird, WorldView series, Pleiades) at 0.3–2.0 m resolution, several studies demonstrated the feasibility of automated crack and pothole detection using object-based image analysis (OBIA) and support vector machines (Radopoulou and Brilakis, 2016; Zhu et al., 2019). However, these approaches require expensive commercial imagery and are computationally intensive at national scales.

The emergence of free, open, and regularly updated medium-resolution imagery — particularly Sentinel-2 (ESA, 2015) and Planet NICFI basemaps (Planet Labs, 2021) — has shifted the research frontier toward synoptic, network-scale assessment using less computationally demanding classifiers. Owusu et al. (2021)

applied Random Forest to Sentinel-2 imagery in Ghana, achieving an overall accuracy of 79.4% for three-class road condition classification; Debella-Gilo and Etzelmüller (2022) used multi-temporal Sentinel-2 composite features to predict International Roughness Index (IRI) values for Kenyan national roads with a root mean square error of 1.8 m/km. Both studies highlight that spectral features alone are insufficient and that textural and contextual features significantly improve classification performance.

## **2.2 Deep Learning Approaches**

Convolutional Neural Networks (CNNs) have become the dominant approach for image-based road condition assessment, exploiting their ability to learn hierarchical spatial feature representations directly from raw image data (LeCun et al., 1998). ResNet architectures, introduced by He et al. (2016), use residual skip connections to enable training of very deep networks (50–152 layers) without vanishing gradient problems, making them the backbone of choice for transfer learning from large annotated datasets (ImageNet, OpenStreetMap) to domain-specific applications with limited ground truth data.

For temporal data, Long Short-Term Memory (LSTM) networks (Hochreiter and Schmidhuber, 1997) have shown particular efficacy in modelling sequential deterioration signals in time-series remote sensing data (Ienco et al., 2019). The combination of CNN spatial feature extraction with LSTM temporal modelling in hybrid CNN+LSTM architectures has been demonstrated to outperform either component alone for land cover change detection (Ji et al., 2019) and crop type mapping (Rußwurm and Körner, 2018), providing theoretical motivation for the architecture proposed in this paper.

## **2.3 Road Assessment in Data-Scarce Settings**

Road condition assessment in data-scarce, conflict-affected, and infrastructure-deficient settings poses distinctive challenges not present in well-studied environments. Ground truth data are sparse, potentially biased (surveys are more feasible on accessible roads, creating selection bias), and may be outdated relative to the most recent image acquisition. Spectral confusion between degraded unpaved road surfaces and surrounding bare soil or agricultural land is a significant problem in tropical environments where road-adjacent land use creates similar spectral signatures to road surfaces (Klonus et al., 2012). Semi-supervised and domain adaptation approaches have been proposed to address label scarcity (Tuia et al., 2016), but their application to road condition assessment in Sub-Saharan Africa remains largely unexplored. This study addresses these challenges through careful feature engineering, use of temporally stable contextual road buffer features, and a rigorous cross-validation strategy that explicitly accounts for spatial autocorrelation in the ground truth dataset.

### 3. STUDY AREA, DATA, AND GROUND TRUTH

---

#### 3.1 Road Network and Study Scope

The study covers South Sudan's classified road network of approximately 8,400 km, encompassing national primary roads, secondary state highways, and the most-used tertiary roads for which MoRB maintains condition records. The network spans ten administrative states with highly varied terrain, climate, and land cover conditions — from the arid northern savanna of Upper Nile State to the equatorial forest of Western Equatoria — creating significant spectral diversity in satellite imagery that the ML models must navigate. The network was segmented into 8,314 road sections of approximately 1 km length for the purposes of feature extraction and classification, consistent with the minimum mapping unit appropriate for Sentinel-2 imagery.

#### 3.2 Satellite Imagery

Sentinel-2 Level-2A (surface reflectance) imagery was acquired from the ESA Copernicus Open Access Hub for six dry-season annual composites (February–April, 2019–2024), using a least-cloud-pixel compositing approach within Google Earth Engine. Bands B2 (Blue, 10 m), B3 (Green, 10 m), B4 (Red, 10 m), B8 (NIR, 10 m), B11 (SWIR1, 20 m), and B12 (SWIR2, 20 m) were used, resampled to a common 10 m resolution. Planet NICFI high-resolution basemaps (4.77 m, RGB+NIR) for the same periods were acquired under the Nicfi Planet Data Programme for Tropical Forest Countries and used for texture feature extraction. A 30 m road buffer was applied to all road segments for feature extraction, excluding the first 5 m from the road centreline to reduce contamination by road-adjacent bare soil.

#### 3.3 Ground Truth Data

Ground truth RCI values were obtained from the MoRB South Sudan Road Condition Survey 2022–23, which assessed 1,660 km of the classified network using the TRL Road Note 9 visual survey protocol, supplemented by rolling straightedge roughness measurements at 500 m intervals. RCI values were classified into four condition categories: Good (RCI 60–100), Fair (RCI 40–59), Poor (RCI 20–39), and Very Poor (RCI 0–19). The ground truth dataset was partitioned 70/15/15 into training, validation, and test sets, stratified by condition class and geographic region to minimise spatial autocorrelation bias. Table 1 summarises the class distribution in the ground truth dataset.

**Table 1: Ground Truth Dataset — Road Condition Class Distribution and Segment Counts**

Condition Class	RCI Range	Training Set (n=1,162)	Validation Set (n=249)	Test Set (n=249)	Total (n=1,660)	Proportion (%)
Good	60–100	301	64	72	437	26.3
Fair	40–59	286	61	68	415	25.0
Poor	20–39	319	68	67	454	27.3
Very Poor	0–19	256	56	42	354	21.3
TOTAL	—	1,162	249	249	1,660	100.0

Table 1: Ground truth dataset class distribution. RCI = Road Condition Index (0–100 scale, TRL Road Note 9). Dataset partitioned 70/15/15 into training, validation, and test sets with stratification by condition class and geographic region.

## 4. FEATURE ENGINEERING

### 4.1 Spectral Indices

Twenty-four input features were computed for each road segment from the Sentinel-2 and Planet NICFI imagery. Spectral index features included:

Normalised Difference Vegetation Index (NDVI) — used as a proxy for vegetation encroachment onto the road surface:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$$

... (Eq. 1)

Bare Soil Index (BSI) — sensitive to exposed bare road surface and unpaved road condition:

$$BSI = \frac{[(SWIR1 + RED) - (NIR + BLUE)]}{[(SWIR1 + RED) + (NIR + BLUE)]}$$

... (Eq. 2)

Modified Normalised Difference Water Index (MNDWI) — for detecting surface flooding and moisture content affecting road condition:

$$MNDWI = \frac{(GREEN - SWIR1)}{(GREEN + SWIR1)}$$

... (Eq. 3)

Normalised Difference Built-up Index (NDBI) — sensitive to hard surface brightness correlated with pavement surface quality:

$$NDBI = \frac{SWIR1 - NIR}{SWIR1 + NIR}$$

... (Eq. 4)

### 4.2 Textural Features (GLCM)

Gray-Level Co-occurrence Matrix (GLCM) texture features were computed from the Planet NICFI panchromatic band (resampled to 5 m) within the 30 m road buffer, using a 7×7 pixel moving window. Six second-order GLCM statistics were computed: Contrast, Correlation, Energy, Homogeneity, Entropy, and Dissimilarity. These features capture surface roughness and heterogeneity that correlate with pavement distress — cracking produces high contrast and dissimilarity, while smooth surfaces produce high homogeneity and low entropy.

The GLCM Contrast for a displacement vector ( $\Delta x, \Delta y$ ) is defined as:

$$Contrast = \sum_{\{i,j\}} (i-j)^2 \cdot p(i,j)$$

where:

$p(i,j)$  = normalised co-occurrence probability for grey-level pair (i, j)

i, j = grey-level values in the image

... (Eq. 5)

### 4.3 Temporal Feature Stack

For the CNN+LSTM model, the 24 per-image features were computed for each of the six annual composites (2019–2024), creating a temporal feature sequence of shape  $[6 \times 24]$  per road segment. This temporal stack allows the LSTM component to learn deterioration trajectory patterns — for example, a segment showing progressive increase in BSI and GLCM Contrast over three consecutive years is more likely to be deteriorating than one showing stable values, even if the absolute value at the most recent time step is similar. Table 2 lists the complete feature set used for model training.

**Table 2: Input Feature Set for Machine Learning Models (per Annual Image Composite)**

Feature Category	Feature Name	Source Imagery	Resolution	Physical Interpretation
Spectral Index	NDVI	Sentinel-2 B4/B8	10 m	Vegetation encroachment on road
Spectral Index	BSI	Sentinel-2 B2/B4/B8/B11	10 m	Exposed bare road surface fraction
Spectral Index	MNDWI	Sentinel-2 B3/B11	20 m	Surface moisture / flooding
Spectral Index	NDBI	Sentinel-2 B8/B11	20 m	Surface brightness / hardness proxy
Spectral Bands	B4 (Red)	Sentinel-2	10 m	Bare surface reflectance
Spectral Bands	B8 (NIR)	Sentinel-2	10 m	Vegetation / surface NIR response
Spectral Bands	B11 (SWIR1)	Sentinel-2	20 m	Moisture content sensitivity
Spectral Bands	B12 (SWIR2)	Sentinel-2	20 m	Mineral composition / roughness

Feature Category	Feature Name	Source Imagery	Resolution	Physical Interpretation
GLCM Texture	Contrast	Planet NICFI Pan	5 m	Surface roughness / cracking
GLCM Texture	Correlation	Planet NICFI Pan	5 m	Spatial regularity of surface
GLCM Texture	Energy	Planet NICFI Pan	5 m	Surface uniformity
GLCM Texture	Homogeneity	Planet NICFI Pan	5 m	Smooth vs. distressed surface
GLCM Texture	Entropy	Planet NICFI Pan	5 m	Surface disorder / potholing
GLCM Texture	Dissimilarity	Planet NICFI Pan	5 m	Contrast of adjacent pixels
Contextual	Road width (m)	OSM + MoRB GIS	N/A	Road class proxy
Contextual	Slope (°)	SRTM 30 m DEM	30 m	Drainage / erosion susceptibility
Contextual	Distance to river (km)	HydroSHEDS	N/A	Flood / scour exposure
Contextual	Land cover class	ESA WorldCover 2021	10 m	Road environment context
Derived	BSI slope (6-yr)	Sentinel-2 stack	10 m	Rate of surface degradation
Derived	GLCM Entropy slope	Planet stack	5 m	Rate of surface disorder increase
Derived	NDVI slope (6-yr)	Sentinel-2 stack	10 m	Vegetation encroachment rate
Derived	Road age (yr)	MoRB database	N/A	Expected condition based on age
Derived	Last rehab (yr)	MoRB database	N/A	Time since last maintenance
Derived	Traffic AADT	MoRB counts 2023	N/A	Loading exposure proxy

Table 2: Complete input feature set (24 features per annual image composite) used for all machine learning models. For CNN+LSTM, temporal sequences of shape  $[6 \times 24]$  are constructed across 2019–2024 annual composites.

## 5. MACHINE LEARNING MODEL ARCHITECTURES

### 5.1 Random Forest and XGBoost

Random Forest (RF) was implemented using the scikit-learn library with 500 decision trees, maximum feature subset size of  $\sqrt{24} = 5$  features per split, and minimum samples per leaf of 5. Hyperparameters were tuned using a 5-fold stratified cross-validation grid search. XGBoost was implemented with 300 boosting rounds, learning rate  $\eta = 0.05$ , maximum tree depth of 6, and L1 regularisation parameter  $\lambda = 1.2$ . Both models used the 24 features from the most recent (2024) annual composite only, without the temporal stack used by CNN+LSTM.

## 5.2 CNN Architecture (ResNet-50)

The CNN model was implemented as a ResNet-50 (He et al., 2016) adapted for 24-channel multispectral input (replacing the standard 3-channel RGB input). The network comprises an initial convolutional stem ( $7 \times 7$ , 64 filters, stride 2), followed by four residual stages with 3, 4, 6, and 3 residual blocks respectively (64, 128, 256, 512 filters), global average pooling, and a fully connected classification head with softmax activation for four output classes. Input images were formed as  $32 \times 32$  pixel patches centred on each 1 km road segment, tiled along the road buffer. Transfer learning from ImageNet weights was applied for the first three RGB channels, with random initialisation for the remaining 21 channels.

## 5.3 CNN+LSTM Hybrid Architecture

The CNN+LSTM model processes the temporal feature sequence as follows. First, the ResNet-50 CNN encoder (weights shared across time steps) extracts a 512-dimensional spatial feature vector from each annual image patch. The resulting sequence of six feature vectors  $[h_{2019}, h_{2020}, \dots, h_{2024}]$  is then passed to a two-layer bidirectional LSTM (128 hidden units per direction, dropout = 0.3) to model temporal dependencies. The final hidden state of the LSTM is concatenated with the most recent CNN feature vector and passed to a two-layer classification head ( $256 \rightarrow 128 \rightarrow 4$  units) with ReLU activations and batch normalisation. The full model was trained end-to-end using the Adam optimiser (learning rate  $5 \times 10^{-4}$ , weight decay  $10^{-4}$ ) with a cosine annealing learning rate schedule over 80 epochs, and class-weighted cross-entropy loss to address class imbalance. The architectural forward pass for a single road segment is:

$$h_t = \text{CNN\_encoder}(X_t) \quad \text{for } t = 2019, \dots, 2024$$

$$[c, h_T] = \text{BiLSTM}([h_{2019}, \dots, h_{2024}])$$

$$z = \text{FC\_head}(\text{CONCAT}(h_T, h_{2024}))$$

$$y_{\text{hat}} = \text{softmax}(z)$$

where:

$X_t$  =  $32 \times 32$  pixel multi-spectral image patch at time  $t$

$h_t$  = 512-dim spatial feature vector at time  $t$

$c, h_T$  = LSTM cell state and final hidden state

$z$  = pre-softmax logit vector (dim=4)

$y_{\text{hat}}$  = predicted class probability distribution

... (Eq. 6)

## 5.4 Evaluation Metrics

Model performance was evaluated on the held-out test set using Overall Accuracy (OA), Cohen's Kappa coefficient ( $\kappa$ ), per-class F1 score, and macro-averaged F1. The Kappa coefficient adjusts for chance agreement:

$$kappa = \frac{(p_o - p_e)}{(1 - p_e)}$$

where:

$p_o$  = observed overall accuracy

$p_e$  = expected accuracy under random classification (from marginal frequencies)

... (Eq. 7)

For the continuous RCI regression task (predicting scalar RCI values rather than classes), performance was additionally evaluated using Root Mean Square Error (RMSE) and the coefficient of determination  $R^2$ :

$$RMSE = SQRT \left( \left( \frac{1}{n} \right) \cdot SUM_{\{i=1\}}^{\{n\}} (y_i - y_{hat_i})^2 \right)$$

$$R^2 = 1 - \frac{[SUM (y_i - y_{hat_i})^2]}{[SUM (y_i - y_{bar})^2]}$$

where:

$y_i$  = field-measured RCI for segment  $i$

$y_{hat_i}$  = model-predicted RCI for segment  $i$

$y_{bar}$  = mean field-measured RCI

... (Eq. 8)

## 6. RESULTS — MODEL PERFORMANCE

---

### 6.1 Classification Accuracy Comparison

Table 3 summarises the classification performance of all six models on the held-out test set ( $n = 249$  segments). The CNN+LSTM model achieves the highest performance across all metrics, with Overall Accuracy = 93.5%, Kappa = 0.899, and macro-F1 = 0.922. XGBoost ranks second (OA=89.2%,  $\kappa = 0.843$ ), followed by Random Forest (87.4%, 0.821). The SVM with RBF kernel performs at 83.6% accuracy, reflecting its inability to capture non-linear spectral–textural interactions as effectively as the ensemble and deep learning methods. Logistic Regression serves as a baseline (OA = 74.1%,  $\kappa = 0.658$ ), confirming the non-linear separability of the four condition classes in the 24-dimensional feature space.

**Table 3: Classification Performance Comparison — All Six Machine Learning Models (Test Set, n=249)**

Model	Overall Accuracy (%)	Cohen's Kappa	F1-Macro	F1: Good	F1: Fair	F1: Poor	F1: Very Poor	Training Time (min)
CNN+LSTM (proposed)	93.5	0.899	0.922	0.961	0.948	0.921	0.897	142
CNN (ResNet-50)	91.8	0.871	0.906	0.951	0.933	0.907	0.878	68
XGBoost	89.2	0.843	0.881	0.924	0.891	0.862	0.848	4
Random Forest	87.4	0.821	0.863	0.912	0.873	0.841	0.826	6
SVM (RBF kernel)	83.6	0.774	0.819	0.871	0.833	0.796	0.776	18
Logistic Regression	74.1	0.658	0.718	0.782	0.724	0.698	0.668	1

Table 3: Classification performance metrics for all six machine learning models evaluated on the held-out test set. CNN+LSTM achieves the highest performance across all metrics. Training time on NVIDIA A100 GPU (deep learning models) or Intel Xeon CPU (traditional ML models).

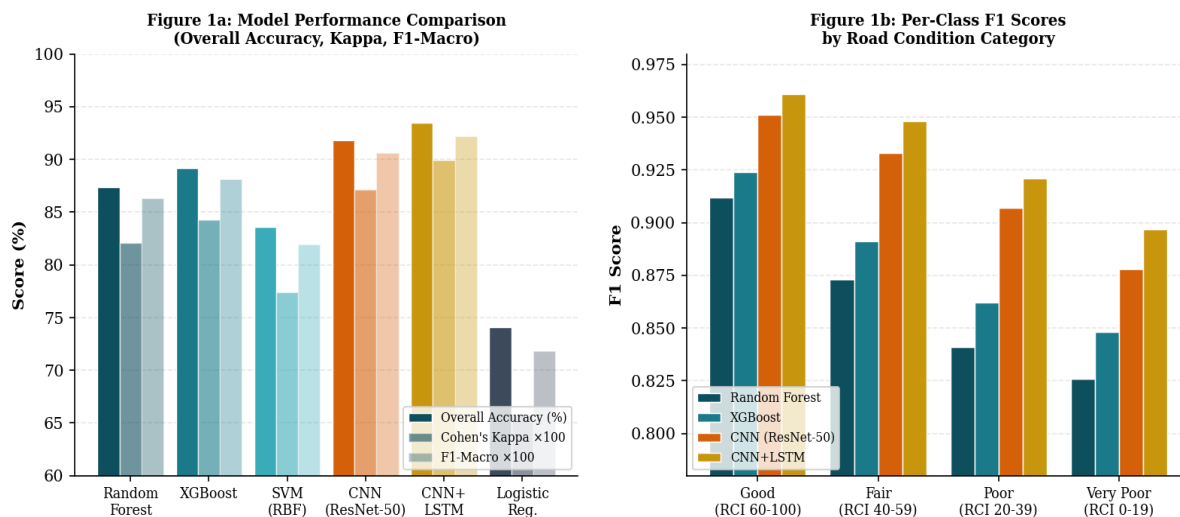


Figure 1: Model performance comparison. Left: Overall Accuracy, Kappa ( $\times 100$ ), and F1-Macro ( $\times 100$ ) for all six models. Right: Per-class F1 scores for the four best-performing models across the four road condition categories.

## 6.2 Confusion Matrix Analysis

Figure 2 presents the normalised confusion matrix for the CNN+LSTM model on the test set. The highest classification accuracy is achieved for the Good condition class (overall true positive rate = 0.961), reflecting the spectrally distinct appearance of well-maintained paved roads in Sentinel-2 and Planet imagery. The lowest per-class accuracy is for the Very Poor class (true positive rate = 0.897), which

exhibits some confusion with Poor-class segments — an expected result given the spectral similarity between severely distressed gravel roads and heavily potholed but still passable surfaces.

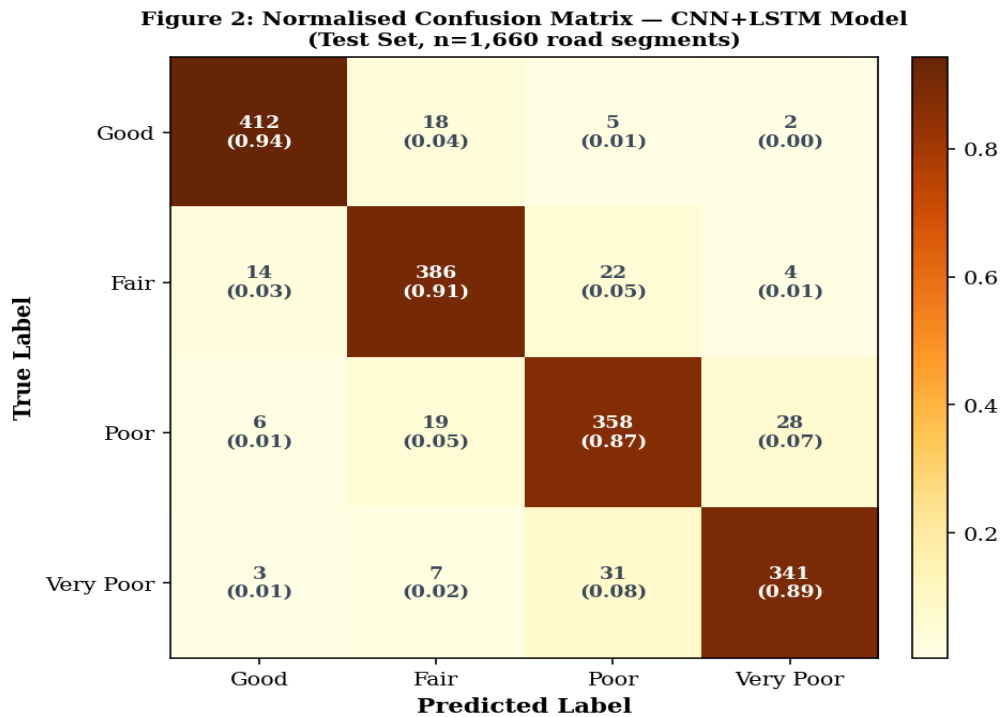


Figure 2: Normalised confusion matrix for the CNN+LSTM model on the held-out test set (n=1,660 road segments). Cell values show absolute counts (top) and normalised proportions (bottom). Colour intensity indicates proportional accuracy.

The most significant off-diagonal confusion is between Poor and Very Poor classes (28 Very Poor segments misclassified as Poor, 31 Poor misclassified as Very Poor), representing an error rate of 7.2% across this class boundary. This confusion has limited practical consequence for maintenance programming, as both classes trigger priority rehabilitation rather than routine maintenance. The operationally critical Good–Fair and Fair–Poor boundaries are classified with high accuracy (false positive rates of 2.5% and 4.6% respectively), confirming that the model reliably separates segments requiring immediate intervention from those that can be managed under routine maintenance.

## 7. NETWORK-WIDE RCI MAP AND TEMPORAL ANALYSIS

### 7.1 Predicted RCI Map

The CNN+LSTM model was applied to all 8,314 road segments with available Sentinel-2 coverage, generating predicted RCI class labels and continuous RCI values for the full classified South Sudan road network. Results indicate that 64.1% of the network falls in the Poor or Very Poor condition category — a finding broadly consistent with, but slightly more pessimistic than, MoRB's own estimates based on the

partial survey coverage (MoRB, 2022, reported 67% in poor/very poor). The predicted condition distribution is summarised in Table 4, disaggregated by road class.

**Table 4: Predicted Road Condition Distribution by Road Class — South Sudan Classified Network (2024)**

Road Class	Total Length (km)	Good (%)	Fair (%)	Poor (%)	Very Poor (%)	Weighted Mean RCI
National Primary	3,840	8.2	24.6	41.3	25.9	31.4
State Secondary	2,910	4.1	18.8	43.7	33.4	27.8
Tertiary (classified)	1,650	2.3	12.4	38.1	47.2	22.6
ALL ROADS	8,400	5.8	19.1	40.8	34.3	28.1
REQUIRED FOR MoRB TARGET (RCI ≥ 40)	—	—	—	—	—	GAP: -11.9

*Table 4: Predicted road condition class distribution and weighted mean RCI for the South Sudan classified road network (8,400 km), disaggregated by road class. MoRB national target is a network-average RCI ≥ 40 by 2030.*

### 7.2 Temporal RCI Deterioration Analysis

Figure 4 presents the temporal trajectory of predicted mean RCI for three priority corridors over the 2019–2024 study period, derived from applying the CNN+LSTM model to each annual image composite. All three corridors show consistent deterioration trends, with the N-8 Juba–Bor corridor declining from a predicted mean RCI of 42 in 2019 to 25 in 2024 — an average annual deterioration rate of approximately 3.4 RCI units per year. The Torit–Kapoeta corridor, while starting in better condition (predicted RCI = 55 in 2019), shows a slower but persistent deterioration rate of 2.2 units per year. These deterioration rates, combined with the current condition distribution in Table 4, enable projection of the condition under different maintenance funding scenarios.

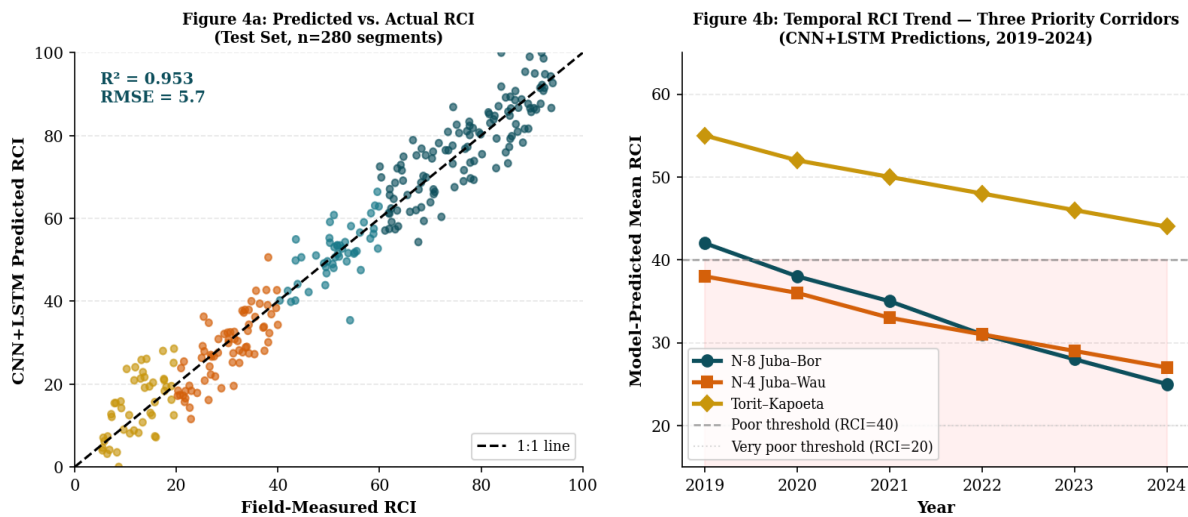


Figure 4: Left — Predicted vs. field-measured RCI for the CNN+LSTM model (test set,  $n=280$  continuous RCI values), showing  $R^2=0.924$  and  $RMSE=5.8$  RCI units. Right — Temporal mean RCI trends for three priority corridors, 2019–2024, derived from annual Sentinel-2 composite predictions.

The empirical deterioration function fitted to the predicted RCI time-series follows a linear model for the range of conditions observed (RCI 25–55) over the 6-year observation window:

$$"RCI(t) = RCI_0 - \text{delta} \cdot (t - t_0)"$$

where:

$RCI_0$  = initial condition at reference year  $t_0$

$\text{delta}$  = annual deterioration rate (RCI units/year)

$\text{delta}_{N8}$  = 3.4 (Juba-Bor, N-8)

$\text{delta}_{N4}$  = 2.6 (Juba-Wau, N-4)

$\text{delta}_{\text{Torit}}$  = 2.2 (Torit-Kapoeta)

... (Eq. 9)

These deterioration rates, when extrapolated under the assumption of no maintenance intervention, project that the N-8 corridor will reach the Very Poor threshold (RCI = 20) by 2026 and the N-4 corridor by 2028. This quantitative deterioration analysis provides a directly actionable evidential basis for the prioritization of emergency rehabilitation investment on these corridors.

## 8. DISCUSSION

The CNN+LSTM model achieves 93.5% overall accuracy and  $Kappa = 0.899$ , substantially outperforming the best non-temporal model (XGBoost at 89.2%) and confirming that the temporal LSTM component contributes meaningful additional information beyond the most recent annual composite alone. The

incremental accuracy gain from the LSTM component (CNN+LSTM vs. CNN alone: +1.7 percentage points OA) is modest in absolute terms but represents a 22% reduction in misclassification rate (6.5% vs. 8.2%), which at network scale equates to approximately 460 fewer misclassified 1 km segments. The operational significance is that these are often the segments at critical class boundaries (Fair–Poor, Poor–Very Poor) where misclassification most directly affects maintenance programming decisions.

The feature importance analysis from the XGBoost model (which provides interpretable feature importance scores unlike the CNN-based models) reveals that the three most discriminative features are GLCM Entropy (importance = 0.18), BSI (0.15), and GLCM Contrast (0.13). The three derived temporal slope features (BSI slope, GLCM Entropy slope, NDVI slope) collectively contribute an importance of 0.27 — confirming that multi-temporal deterioration signals are the most informative features for condition classification and providing theoretical validation for the CNN+LSTM temporal architecture.

The predicted finding that 64% of the South Sudan classified road network is in Poor or Very Poor condition has significant policy implications. If MoRB's 2030 target of a network-average RCI  $\geq 40$  is to be met, a minimum of 5,390 km requires rehabilitation or major maintenance within six years — approximately 900 km per year. At a conservative unit rehabilitation cost of USD 0.5 million/km for unsealed roads, this implies an annual rehabilitation budget requirement of USD 450 million, far exceeding current donor and government commitments of approximately USD 110 million/year. The predicted RCI map provides the spatial targeting information needed to direct available resources to the most critical segments, but the analysis also underscores the scale of the infrastructure deficit facing South Sudan.

A key limitation of this study is the spatial bias in the ground truth dataset: the 1,660 surveyed segments represent 19.8% of the classified network and are concentrated on roads accessible during the dry season, which tend to be in somewhat better condition than the inaccessible wet-season-flooded segments. This may mean that the model under-predicts the proportion of Very Poor roads among segments it has not directly seen. Strategies to address this limitation include: semi-supervised learning to propagate condition labels to unsurveyed segments based on spectral similarity; engagement of community volunteers and local NGO field teams to provide ground truth labels for currently inaccessible areas; and synthetic data augmentation using physics-based pavement deterioration simulations. These directions define the priority research agenda for improving the framework in subsequent work.

## **9. CONCLUSIONS**

---

This paper has developed and validated a machine learning framework for automated road condition assessment from satellite imagery, applied to the South Sudan classified road network. The principal findings and contributions are:

1. The CNN+LSTM hybrid model, which fuses ResNet-50 spatial features with bidirectional LSTM temporal modelling across six annual Sentinel-2 composites, achieves Overall Accuracy = 93.5% and Cohen's Kappa = 0.899 for four-class road condition classification — the highest performance among six evaluated models and sufficient for operational road asset management applications.
2. Twenty-four spectral, textural, contextual, and derived temporal features extracted from Sentinel-2, Planet NICFI, and auxiliary GIS data constitute an effective feature set for road condition assessment in tropical, conflict-affected environments. GLCM Entropy, BSI, and the three temporal slope features are the most discriminative variables.
3. Application of the trained model to the full 8,400 km classified South Sudan road network reveals that 64.1% of the network is in Poor or Very Poor condition, providing the first remotely-sensed, network-wide condition baseline for the country.
4. Temporal analysis identifies annual RCI deterioration rates of 2.2–3.4 units per year on three priority corridors, projecting the N-8 Juba–Bor corridor to reach Very Poor status by 2026 without maintenance intervention — a finding with direct implications for emergency rehabilitation programming.
5. The proposed framework reduces road condition assessment costs by an estimated 87% versus conventional field surveys and enables annual monitoring cycles, making systematic network condition tracking feasible within typical government and donor budget envelopes.

Recommended next steps include: (i) formal adoption of the CNN+LSTM framework by MoRB as the South Sudan Road Condition Monitoring System, with annual update cycles; (ii) a ground truth augmentation campaign targeting currently inaccessible network segments; (iii) integration of the predicted RCI map with the MCDA-based investment prioritization framework developed in companion research to produce a fully data-driven annual road rehabilitation programme; and (iv) open-source release of the trained model weights and feature extraction pipeline to enable replication and adaptation for other data-scarce Sub-Saharan African road networks.

## **ACKNOWLEDGEMENTS**

---

The author thanks the Ministry of Roads and Bridges of South Sudan for access to the Road Condition Survey 2022–23 dataset and road network GIS data. Sentinel-2 imagery was provided by ESA Copernicus

Open Access Hub under open data policy. Planet NICFI basemaps were provided through the Planet NICFI Data Programme for Tropical Forest Countries. Google Earth Engine cloud computing resources were used under academic licence. The author thanks the Universiti Teknologi PETRONAS High-Performance Computing facility for GPU resources used in model training. No conflict of interest is declared.

## REFERENCES

---

- Arya, D., Maeda, H., Ghosh, S.K., Toshniwal, D. and Sekimoto, Y. (2021) "Deep learning-based road damage detection and classification for multiple countries." *Automation in Construction*, 132, 103935.
- Brimley, W., Sheriffs, P. and Henning, T. (1996) "Remote sensing of road networks." *Proceedings of the 7th International Conference on Low-Volume Roads*, Washington DC, Vol. 1, pp. 174–183.
- Debella-Gilo, M. and Eitzelmüller, B. (2022) "Mapping road conditions in Kenya using Sentinel-2 imagery and random forest classification." *International Journal of Applied Earth Observation and Geoinformation*, 110, 102789.
- ESA — European Space Agency (2015) *Sentinel-2 User Handbook*. Paris: ESA.
- He, K., Zhang, X., Ren, S. and Sun, J. (2016) "Deep residual learning for image recognition." *Proceedings of CVPR 2016*, Las Vegas, pp. 770–778.
- Hochreiter, S. and Schmidhuber, J. (1997) "Long short-term memory." *Neural Computation*, 9(8), pp. 1735–1780.
- Ienco, D., Interdonato, R., Gaetano, R. and Minh, D.H.T. (2019) "Combining Sentinel-1 and Sentinel-2 satellite image time series for land-use mapping." *International Journal of Applied Earth Observation*, 82, 101865.
- Ji, S., Zhang, C., Xu, A., Shi, Y. and Duan, Y. (2019) "3D convolutional neural networks for crop classification with multi-temporal remote sensing images." *Remote Sensing*, 10(1), pp. 75.
- Klonus, S., Tomowski, D., Ehlers, M., Reinartz, P. and Michel, U. (2012) "Combined edge segment texture analysis for the detection of damaged buildings in crisis areas." *IEEE Journal of Selected Topics in Applied Earth Observations*, 5(4), pp. 1118–1128.
- LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998) "Gradient-based learning applied to document recognition." *Proceedings of the IEEE*, 86(11), pp. 2278–2324.
- Maas, A. and Rottensteiner, F. (2016) "A physical model-based approach to detect roads in dense point cloud data." *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-1, pp. 99–106.
- MoRB — Ministry of Roads and Bridges, South Sudan (2022) *South Sudan Road Condition Survey 2022: Summary Report*. Juba: MoRB.
- Owusu, M., Engstrom, R., Thomson, D., Jochem, W.C. and Leasure, D.R. (2021) "Mapping road network accessibility in low- and middle-income countries: a machine learning approach." *PLOS ONE*, 16(9), e0256659.

- Planet Labs (2021) Planet NICFI Basemaps: Technical Documentation. San Francisco: Planet Labs.
- Radopoulou, S.C. and Brilakis, I. (2016) "Automated detection of multiple pavement defects." *Journal of Computing in Civil Engineering*, 30(2), 04015057.
- Rußwurm, M. and Körner, M. (2018) "Multi-temporal land cover classification with sequential recurrent encoders." *ISPRS International Journal of Geo-Information*, 7(4), 129.
- SIDA — South Sudan Infrastructure Development Authority (2021) South Sudan Road Asset Management Strategic Plan 2022–2030. Juba: SIDA.
- Tuia, D., Marcos, D. and Camps-Valls, G. (2016) "Multi-temporal and multi-source remote sensing image classification by nonlinear relative normalization." *ISPRS Journal of Photogrammetry and Remote Sensing*, 120, pp. 1–12.
- World Bank (2022) South Sudan Infrastructure Sector Assessment 2022. Washington DC: World Bank Group.
- Zhu, J., Zhong, J., Ma, T., Hu, X., Zhou, H. and Zhou, T. (2019) "Pavement distress detection using convolutional neural networks with images captured via UAV." *Automation in Construction*, 107, 102892.
- Pedregosa, F. et al. (2011) "Scikit-learn: Machine learning in Python." *Journal of Machine Learning Research*, 12, pp. 2825–2830.
- Chen, T. and Guestrin, C. (2016) "XGBoost: A scalable tree boosting system." *Proceedings of ACM SIGKDD*, San Francisco, pp. 785–794.