

# Deep Learning Inter-city Road Conditions in East Africa Focusing on Rwanda for Infrastructure Prioritization using Satellite Imagery and Mobile Data

Davy K. Uwizera<sup>1</sup>, Charles Ruranga<sup>2</sup>, and Patrick McSharry<sup>3</sup>, *Fellow, RSS, Senior Member, IEEE*

**Abstract**—Traditional survey methods for gathering information, such as questionnaires and field visits, have long been used in East Africa to evaluate road conditions and prioritize their development. These surveys are time-consuming, expensive, and vulnerable to human error. Road building and maintenance, on the other hand, has long experienced multiple challenges due to a lack of accountability and validation of conventional approaches to determining which areas to prioritize. With the digital revolution, a lot of data is generated daily such as call detail record (CDR), that is likely to contain useful proxy data for spatial mobility distribution across different routes. In this research we focus on satellite imagery data with applications in East Africa and Google Maps suggested inter-city roads to assess road conditions and provide an approach for infrastructure prioritization given mobility patterns between cities. With increased urban population, East African cities have been expanding in multiple directions affecting the overall distribution of residential areas and consequently likely to impact the mobility trends across cities. We introduce a novel approach for infrastructure prioritization using deep learning and big data analytics. We apply deep learning to satellite imagery, to assess road conditions by area and big data analytics to CDR data, to rank which ones could be prioritized for construction given mobility trends. Among deep learning models considered for roads condition classification, EfficientNet-B3 outperforms them and achieves accuracy of 99%.

**Index Terms**—Deep-learning, mobile data, classification, vision recognition, big data, satellite imagery

## I. INTRODUCTION

### A. Background

INTERCITY roads have always been important to our countries' development because they enable capital and labor to move easily through regions. Despite recent efforts to improve inter-city routes to promote trade of goods across East African cities, a significant number of roads directly connecting cities and districts remain in poor condition [22]. Some of these routes are vulnerable to bad weather due to a lack of pavement, making them dangerous to use, especially during rainy seasons [24].

While there are technologies which address routing such as Google Maps, they tend to focus on the fastest route. And

options they have for avoiding problematic road conditions such as tolls and railways, may not be applicable in developing countries[3]. Furthermore, in East Africa and most developing countries, the methodology and process for assessing road conditions continues to rely heavily on field visits, surveys, and questionnaires, which take a long time [3], are expensive, and are vulnerable to human error [7]. And as a result of the low visibility of these methods, corruption can occur in some cases [27][10].

However, with the rise of digital devices and access to continuously updated satellite imagery, it is possible to support or replace survey based methods with new methods that could enable conducting the assessment of road conditions and infrastructure prioritization more quickly and accurately [30]. This would provide more visibility into road infrastructure budgeting, enhance mobility and reduce travel risk under unfavorable weather conditions. And given the fact that weather patterns have been changing in various parts of the world due to climate change [21], better ways of understanding road conditions could also be very instrumental to transport industry for adaptation planning [9][4].

Other than travel risk, there is also a number of not-well connected routes between neighboring cities, whereby traveling between neighboring cities is only easier if you have to use a third connecting district. These cases are known for hindering economic activities across neighboring cities affecting various sectors such as agriculture, tourism, education, health care etc, where there is always a constant need of moving products or capital from one side to another. Thus, a good way of assessing badly connected city routes would also further contribute to better prioritization of route maintenance and construction across regions for the benefit of various sectors.

We assess route conditions using deep learning methods and satellite imagery and provide a data oriented approach to infrastructure prioritization, using mobility trends extracted from fully anonymized Call Detail Record (CDR).

While machine learning techniques and deep learning methods in particular are now viewed as state-of-art in various fields, their application still lags behind in developing countries, especially in the field of remote sensing [34]. Consequently, some of the machine learning techniques such as deep-learning, which have demonstrated great potential for the analysis and modeling of unstructured data such as imagery,

<sup>123</sup>All authors are with the African Center of Excellence in Data Science (ACE-DS), University of Rwanda. <sup>1</sup>Davy K. Uwizera is an alumnus of Carnegie Mellon University and currently pursuing a dual PhD in Data Science at ACE-DS and Internet of Things at ACEIoT. <sup>2</sup>Dr. Charles Ruranga is currently the director of ACE-DS. <sup>3</sup>Patrick McSharry is a professor of machine learning and big data at Carnegie Mellon University and Oxford University

are not widely used in infrastructure related studies in East Africa. This research also seeks to fill this gap by contributing to existing academic knowledge of applying machine-learning techniques to solve real life problems in East Africa using deep learning[12][36] and big data analytics [37][23].

There are a number research works that have been done in line with route conditions assessment [16][2][28], however most of them are not applicable to Africa context, and a few, none to our knowledge, have considered using both CDR based mobility trends and road conditions to provide an infrastructure prioritization approach in East Africa.

In this paper we conduct our research in East Africa, focusing mainly on Rwanda, located at latitude:-1.940278 and longitude:29.873888. We assess 133 inter-city connecting routes suggested by Google Maps API in the country and provide a way for quick road condition classification and infrastructure prioritization using satellite imagery and CDR data.

## B. Literature Review

Route infrastructure analysis and classification has been of great interest to the research community around the world especially for infrastructure monitoring and policy making. Given a strong connection of road infrastructure and other economic sectors, patterns underlying efficient and bad road networks and their impact on other sectors, have been studied in addressing multiple research problems across the globe.

Gibbons et al. assessed the impact of construction of new road infrastructure in small towns in Britain. The researchers measured exposure to transport improvement and estimated its impact on employment and labor productivity. The research shows that improved transport infrastructure such as construction of new road schemes has substantial positive effects on area level employment and number of establishments. Where by 1% increase to good road accessibility led to 0.3%-0.5% increase in establishments and employment. The paper also states that there was a strong positive correlation between improved transport infrastructure and output per worker, wages and use of intermediate inputs [11].

Petrucchi et al. assessed road conditions during bad weather periods in Italy. Researchers highlight that during bad weather conditions, roads with bad conditions are likely to be affected and cause life threatening risk. Landslides were the most frequent cases. Among the assessed cases 84% of people who were injured were motorists. And of the victims, 50 % were killed along fast-flowing roads [26]. This highlights the importance of well designed routes and continuous assessment of road conditions.

Oshri et al. used deep learning and satellite imagery to assess infrastructure quality in Africa [25]. While infrastructure quality is one of the important features in United Nations Sustainable Development Goals, the research highlights how monitoring progress in developing regions is still expensive using traditional survey methods and proposed a remote sensing approach using deep learning. By training convolutional neural networks (CNN), researchers were able to predict the ground truth labels of the Afrobarometer round five survey.

The labels comprised Electricity, Sewerage, Piped Water and Roads. The researchers achieved classification accuracies of 88%, 86%, 74% and 78% respectively.

Berriel et al. used deep learning for automatic classification of crosswalks [5]. The research focuses on filling the gap of zebra crossing classification problem for automatic retrieval and annotation of crosswalks satellite imagery data. The results of the research showed that deep learning could be used to perform crosswalk classification with a 97% accuracy on a global scale.

Surface object segmentation has also been done using satellite imagery and deep learning, demonstrating the potential of deep learning models to not only identify, but also detect the boundaries of the object pattern throughout the learning process. Using deep learning and satellite imagery, Bischke et al. segmented building footprints [6]. The study centered on using a multi-task loss to develop semantic segmentation approaches. The results showed that they were able to boost accuracy by eight percent over current state-of-the-art semantic segmentation methods.

Khan et al. studied automatic target object detection in satellite images [18]. According to the writers, conventional image feature extraction methods including Histogram of Oriented Gradients, Gabor Feature, and Hough Transformation do not work well for large datasets with varying sizes, orientations, and noisy backgrounds. They propose using CNN to process the set of target object pixels positions using an edge boxes method. Van also used YOLT (you just look twice) CNN architecture to identify small and large objects in satellite imagery, such as vehicles, planes, boats, buildings, and airports, more quickly [35].

There are various ways of analyzing and classifying infrastructure to improve upon the traditional use of surveys.

Cole et al. conducted a research on the analysis and detection of power outages in urban areas that may appear as a result of natural and anthropogenic hazards damaging physical infrastructure. To address this researchers applied a feed forward neural network model on night light satellite data and locally-relevant geo-spatial data, achieving Pearson coefficients between 0.48 and 0.58. Cadamuro et al. assess road infrastructure quality in Kenya using satellite imagery. The researchers demonstrate how deep learning methods on satellite imagery could be used for assigning grades on road infrastructure in developing countries. In this paper the grades considered were great, good, fair, bad, poor [8].

Khalil et al. assessed infrastructure growth in Dubai using multi-temporal satellite imagery. The researchers use supervised and unsupervised learning classification methods such as Jenks Natural Breaks Classification method, with image processing to detect locations which experienced massive growth of infrastructure in the city [17]. The results of the research show that one could confidently quantify infrastructure growth using satellite imagery and machine learning.

Clearly, a great deal of research has gone into using deep learning on satellite imagery, and the findings have been remarkable as compared to conventional image classification and detection methods. In conclusion, as compared to conventional field-intensive survey approaches, the combination of deep

learning and satellite imagery has proven to be very cost-effective and less time-consuming.

In this research we seek to use deep learning methods for classification of roads in East Africa. The four categories considered are (1) paved big road; (2) paved small road; (3) not-paved big; (4) not-paved small (5) industrial zone. In the following section we discuss our methodology starting with the exploration of our data. We also use CDR data to assess infrastructure prioritization for not-paved roads given mobility patterns.

## II. METHODOLOGY

### A. Data

Satellite imagery used in this research is retrieved from Google earth which is supported by Data SIO (Scripps Institution of Oceanography), NOAA (National Oceanic and Atmospheric Administration), US. Navy (United States Navy), NGA (National Geospatial-Intelligence Agency), GEBCO (General Bathymetric Chart of the Oceans), Image Landsat, and Image IBCAO (International Bathymetric Chart of the Arctic Ocean). For resolution, we used a 1000 meters eye altitude and a 400 by 400 view per selection.

For mobility trend study, we have used **fully anonymized CDR data** collected by Rwanda's major network operators (MNOs).

Below we describe and visualize various metrics of the datasets being considered.

### B. Roads Data

Using random sampling of spatial route areas, 50,000 thousands images were retrieved under the four road categories: paved big (5000), paved small (5000), not-paved big (10000), not-paved small, (20000). Our target areas were guided by field survey results obtained from Rwanda Transport Development Agency (RTDA) report [1] on spatial distribution of inter-city routes in Rwanda as highlighted in the map below, Figure 1.

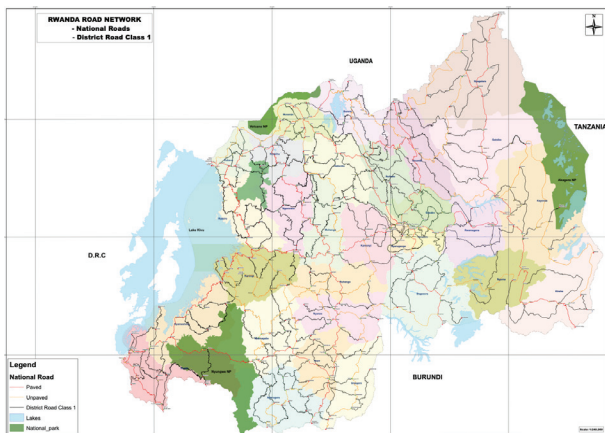


Figure 1: Source: RTDA, Spatial distribution of routes and conditions in Rwanda, 2020.

Using human review and spatial distribution of roads per area as displayed in Figure 1, and their associated legend,

images which were labeled as paved-big, Figure 2, were district class road 1. Most of these consisted of big roads within districts.



Figure 2: District big paved roads sample in Rwanda

Labels for paved-small were given to areas marked as only paved in Figure 1, which consisted of mainly small sized paved roads, Figure 3.

Images which were labeled not-paved-big were roads labeled as unpaved in Figure 4, but which had bigger width size than paved-small class, Figure 3.

Unpaved roads with same or less than the size of paved-small roads class, Figure 3, were marked as unpaved-small, Figure 5.

### C. CDR Data

We analyze 3,01 billion of **fully anonymized** CDR events in March, 2020 with a total of 17 million subscribers. It is towards the end of this month that the first lock-down was introduced in the country due to the pandemic, Covid19, on 21 of March. Hence the insights from this month will capture both before and after lockdown announcement mobility patterns. CDR data is comprised of many fields, but our field of interest for this research were, msisdn for subscriber ID, tower ID from which the event happened, tower ID and it's latitude and longitude, and date-of-charge timestamp which is basically the timestamp at which the CDR event was charged by the MNOs

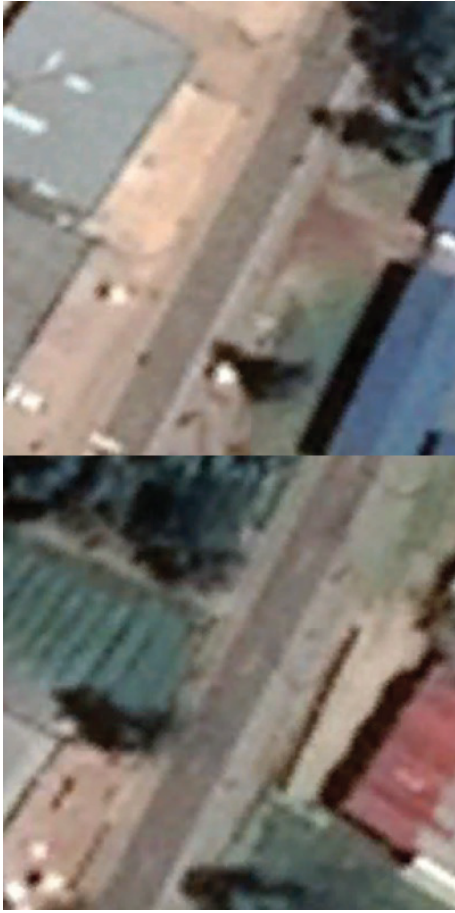


Figure 3: District small paved roads sample in Rwanda



Figure 4: District big unpaved roads sample in Rwanda

from subscriber's account. In Figure-6,7, we visualize CDR events throughout the considered month of March. The results in 7 show that there were more cdr events produced after the announcement of first lock-down due to Covid-19, perhaps as a result of the cashless oriented policies that followed the announcement in favor of mobile transactions.

#### D. Retrieving Inter-districts Routes

To focus on inter-city connecting routes, we use Google Maps APIs to retrieve route positions for all pairs of districts with associated travel time and distance. Travel time and distance retrieved from Google Maps, together with classification of road conditions using satellite imagery, are used in the localization of not-well connected districts routes. See section II-G, II-C. They were also used in mining CDR data, particularly in checking long trips subscribers have done throughout the day and filter out noise in CDR data. Filtering is done by checking one's traveling speed between two locations. Trips which were above 100KM/h were removed from consideration.

#### E. Deep Learning

Road conditions are classified using deep learning methods into four categories: paved-big, paved-small, not-paved-big, and not-paved-small. For training, we use MobileNet version

2 (Mobilenet-V2), Inception version 3 (Inceptionv3), Resnet50 version 2 (Resnet50V2), and EfficientNet (EfficientNet-B3) deep neural network architectures. We also adopt deep transfer learning [31] with imagenet pre-trained weights.

1) *InceptionV3*: The model architecture aims to increase accuracy while minimizing added computation per layer by using factorized convolutional layers and aggressive regularization. InceptionV3 architecture is implemented as in Szegedy et al. [29].

2) *ResNet50V2*: Resnet50V2 is among deep residual network architectures, they focus on learning residual functions rather than learning unreferenced functions. Resnet architectures demonstrated higher accuracy and fast convergence[60]. Resnets models fall in different categories depending on the number of layers used in model building. The most common are Resnet18, Resnet34, Resnet50, Resnet101 and Resnet153 [13]. In this research we use Resnet50V2 [14]. Resnets add non-linearity between layers and the main difference between Resnet version 1 and Resnet version 2 is the removal of non-linearity in the last layer on version 2. This was done to promote easy continuous connection in a form of identity connections. Resnet architecture used are presented in He et al. [14].

3) *EfficientNet-B3*: EfficientNets focus on efficient scaling of deep learning modeling [32]. While some of the mod-



Figure 5: District small unpaved roads sample in Rwanda

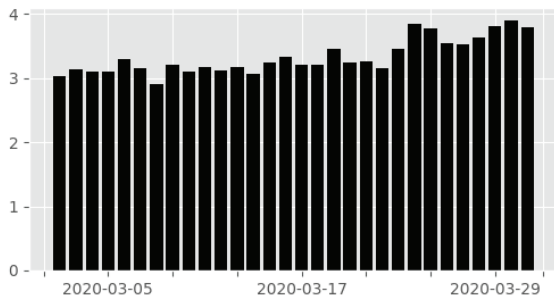


Figure 6: Daily estimated CDR % of total events - March, 2020, Rwanda

els have interchangeably focused on depth and width based scaling. EfficientNet uniformly scales each dimension with a fixed set of scaling coefficients [15]. The proposed method of scaling coefficients in EfficientNets was proved to outperform conventional methods which use arbitrarily scaling of network dimensions, such as focusing only on width, depth and resolution [32]. There are multiple structures of EfficientNets, in this research we used EfficientNet-B3.

As previously mentioned, InceptionV3, Resnet50V2 and EfficientNet-B3 are used as our base-models for our research

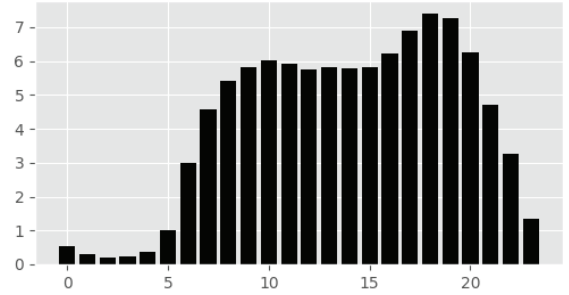


Figure 7: Average estimated % CDR events per hour of the day - March, 2020, Rwanda

task. The models are initiated without the final layer (head) and imagenet weights are used with these models to speed up learning and minimize risk of convergence to local minimum using transfer learning. Then we add one dense layer of size 512 with sigmoid activation and the output layer with softmax activation.

4) *Compactness and Separability Visualization*: To visually understand the separability and compactness of considered predictive models, we use t-stochastic neighborhood embeddings (t-SNE) [19]. t-SNE converts similarities between data points to joint probabilities.

#### F. Origin Destination Matrices

There are different ways for computing OD matrices. Some approaches consider computing home location, where home location is defined as the position where you were most spotted in specified period between 7PM and Midnight, and then joining one's home-location against every other location they are spotted in other than their home-location. This is done sequentially during the day.

Others match one's previously spotted location from the start of day against subsequent locations. Both scenarios could have more complex variants depending on the target goal of the OD matrix [33]. But in most cases, the end goal is to maximize accuracy of the estimated travel location demand, and make sure to filter out data points that could introduce noise in the dataset. Most of the noise in the case of CDR data, comes in when subscribers produce CDR activities during travel not necessarily because they have stopped by that particular location where the activities happened.

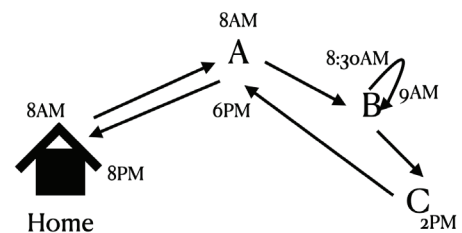


Figure 8: CDR-based Mobility pattern illustration example

To address this for inter-city trips extraction, We divide the day in three window times, midnight to 10am as window 1, 10am to 4pm, window 2 and 4pm to midnight, window 3, Figure 8, and extract OD matrices based on these three windows.

#### G. Localization of not-well connected Neighboring Districts Routes

To identify neighboring districts with not-well connected routes, we extract travel time between neighboring districts from Google Maps APIs and find a trio districts (ABC) where travel time  $t_{AB} - t_{AC} + t_{CB} \leq t$ .

Implying that it takes longer to travel directly from A to B than traveling from A through C to B. Cases where this is true, districts A to B, are likely to have no direct connecting routes or just have bad designed connecting routes, Algorithm-1.

---

#### Algorithm 1 Not-well Connected Neighboring Districts Routes based on Travel Time

---

**Require:**  $f(x, y, z) = T(x, y) + T(y, z)$  where  $T(x, y)$  is the travel time from point  $x$  to  $y$  and  $T(x, y)$  is the travel time from point  $y$  to  $z$  where  $x, y, z$  in this case are districts.  $R$  is the set of considered routes.

**Ensure:**  $C(x, y) = \text{True}$  if  $x$  and  $y$  are neighbors else  $C(x, y) = \text{False}$ .  $R_{i,x,y} = T(x, y)$ , for route  $i$ ,  $i \leq N$  where  $N$  is the total number of routes considered,  $R$ .

$N \leftarrow \text{length}(R)$

$C(x, y) \leftarrow \text{True}$

$C(x, z) \leftarrow \text{False}$

**for**  $i \leftarrow 1$  to  $N$  **do**

$t_{ABC} \leftarrow f(x, y, z) \leftarrow T(x, y) + T(y, z)$

$t_{ACB} \leftarrow f(x, z, y) \leftarrow T(x, z) + T(z, y)$

**if**  $t_{ABC} > t_{ACB}$  **then**

$R_{i,x,y} = \text{not-well connected}$

**end if**

---

Besides, neighboring districts connected with a not-paved road are automatically included among districts with not-well connected routes, Algorithm-2.. We use deep learning methods previously discussed in section II-E, to identify pairs of districts with not-paved connecting routes.

#### H. Route Infrastructure Prioritization

Given not-well connected districts identified using methodology in section II-G, let total of movements between a pair of not-well connected districts, AB, over total movement between neighboring districts be MAB, total movement in and out A be MA, total movements in and out B be MB.

Let average distance between A to 3 most travel location cities be AVDA and average distance between B to 3 most travel location cities be AVDB.

Let total movements in and out of districts be TM, total movement between pairs of neighborign districts be TNDM, social economic index (SEI) and Financial Economic Index (FEI).

There are multiple approaches of computing SEI and FEI, but as indicated by Marcelo et al. [20], SEI, Formula-1, is composed mainly of the following indicators: direct jobs created

---

#### Algorithm 2 Not-well Connected Neighboring Districts Routes based on Travel Time and Not-paved Condition

---

**Require:**  $D_X = g(f(x)) = g(W^T X + B) = g(w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b)$  where  $g(x)$  is an activation function and  $g(f(x))$  a fine-tuned baseline deep model,  $n$ , the number of layers, and  $X$  the input, which in this case is the road condition image.  $f(x, y, z) = T(x, y) + T(y, z)$  where  $T(x, y)$  is the travel time from point  $x$  to  $y$  and  $T(x, y)$  is the travel time from point  $y$  to  $z$  where  $x, y, z$  in this case are districts.  $R$  is the set of considered routes.

**Ensure:**  $C(x, y) = \text{True}$  if  $x$  and  $y$  are neighbors else  $C(x, y) = \text{False}$ .  $R_{i,x,y} = T(x, y)$ , for route  $i$ ,  $i \leq N$  where  $N$  is the total number of routes considered,  $R$ .

$N \leftarrow \text{length}(R)$

$C(x, y) \leftarrow \text{True}$

$C(x, z) \leftarrow \text{False}$

$D_X > t \leftarrow \text{not-paved} \quad \triangleright t$  is the deep classification threshold

**for**  $i \leftarrow 1$  to  $N$  **do**

$t_{ABC} \leftarrow f(x, y, z) \leftarrow T(x, y) + T(y, z)$

$t_{ACB} \leftarrow f(x, z, y) \leftarrow T(x, z) + T(z, y)$

**if**  $t_{ABC} > t_{ACB}$  or  $D_X > t$  **then**

$R_{i,x,y} = \text{not-well connected}$

**end if**

---

(DJC), number of direct beneficiaries (NDB), people affected by land use restructuring (PA), possible cultural and environmental risks and pollution in terms of CO2. On the other hand FEI, Formula-2, mainly consist of financial internal rate of return (RR), multiplier effects (ME), Priority Economic Zones (PE), Implementation Risk (IR), Complementary/Competition effects (CC). For more info on SEI and FEI implementation see Marcelo et al. [20].

$$SEI = w1DJC + w2NDB + w3PA + w4CO2 + e_i \quad (1)$$

$$FEI = w1RR + w2ME + w3PE + w4AIR + w5CC + e_i \quad (2)$$

While most of researches focus mainly on SEI and FEI to build infrastructure prioritization score of which variables, in most cases, consist of qualitative variables [20], we propose a novel approach for inter-city route infrastructure prioritization which introduces a quantitative component that takes into consideration movements in and out, visual road condition, travel time distance between adjacent cities, and their distances from main cities such as capital cities.

We compute route prioritization score (RPS) for route AB as,

$$RPS_{(A,B)} = b1 \frac{\frac{MA}{TM} + \frac{MB}{TM} + \frac{MAB}{TNDM}}{AVDA^c + AVDB^c} + b2P + b3SEI + b4FEI \quad (3)$$

Whereby coefficient  $c$ , is the distance disadvantage from famous travel location cities and  $b$  are weights to considered variables.  $P$  is an output of 0 if road paved or 1 if not paved, derived from a classification deep learning model,  $D_X$

$= g(f(x)) = g(W^T X + B) = g(w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b)$  where  $g(x)$  is an activation function and  $g(f(x))$  a fine-tuned baseline deep model,  $n$ , the number of layers, and  $X$  the input to the model, which in this case is the road condition image.

For the scope of this research we keep  $b1 = 1$ ,  $b2 = 0$ , 1 and  $b3 = b4 = 0$ . With  $b2 = 1$ , non-paved roads are given high priority, whereas for  $b2 = 0$ , both paved and non-paved roads are given equal priority.

To allow diversification while setting the 3 most travel locations, cities in the capital city are considered as one city. Since cities in Kigali are combined on AVDA and AVDB, we take the shortest distance to one of the three cities in Kigali. Algorithm-3, highlights the flow of the proposed approach.

---

**Algorithm 3** Intercity Connecting Routes Infrastructure Prioritization

---

**Require:**  $D_X = g(f(x)) = g(W^T X + B) = g(w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b)$  where  $g(x)$  is an activation function and  $g(f(x))$  a fine-tuned baseline deep model,  $n$ , the number of layers, and  $X$  the input, which in this case is the road condition image.  $f(x, y, z) = T(x, y) + T(y, z)$  where  $T(x, y)$  is the travel time from point  $x$  to  $y$  and  $T(x, y)$  is the travel time from point  $y$  to  $z$  where  $x, y, z$  in this case are districts.  $R$  is the set of considered routes.

**Ensure:**  $C(x, y) = \text{True}$  if  $x$  and  $y$  are neighbors else  $C(x, y) = \text{False}$ .  $R_{i,x,y} = T(x, y)$ , for route  $i$ ,  $i \leq N$  where  $N$  is the total number of routes considered,  $R$ .

$N \leftarrow \text{length}(R)$

$C(x, y) \leftarrow \text{True}$

$C(x, z) \leftarrow \text{False}$

$D_X > t \leftarrow \text{not-paved} \quad \triangleright t$  is the deep classification threshold

**for**  $i \leftarrow 1$  to  $N$  **do**

$ABC \leftarrow f(x, y, z) \leftarrow T(x, y) + T(y, z)$

$ACB \leftarrow f(x, z, y) \leftarrow T(x, z) + T(z, y)$

**if**  $ABC > ACB$  or  $D_X > t$  **then**

$\text{score} \leftarrow RPS(x, y) \quad \triangleright \text{Formula-3}$

**end if**

---

### I. Testing

The dataset, as described in II-A subsection, is divided into training and testing dataset. 70% is used for training and 30% is used for testing for our target models. The test results are then used to evaluate both accuracy and speed in order to draw conclusions about the best model to recommend for policymakers and practitioners.

## III. RESULTS & DISCUSSION

### A. Modeling

Each model was trained on batch size 32, and 16 epochs with weight decay of 5e-04 and the overall accuracy of EfficientNet-B3, InceptionV3 and Resnet50 is 93%, 95%, 98% respectively. Input shape used was 224x224 for EfficientNet-B3 and Resnet50, 299x299 for Inception, 300x300 for EfficientNet-B3. In the following parts of this section we discuss in detail the classification results of our three target models.

1) *InceptionV3*: Below we present detailed classification results of InceptionV3, Table-I.

Table I: InceptionV3 classification results

	precision	recall	f1-score	Support
Paved-big	1.00	0.94	0.97	1500
Paved-small	0.89	0.97	0.93	1500
Not-paved-big	0.94	0.92	0.93	3000
Not-paved-small	0.94	0.96	0.95	6000
Accuracy				0.95

Inception scores 95% overall accuracy on the four road condition categories. Paved-big, Not-paved-small and Not-paved small areas are the most well classified categories by the model. Table-I.

2) *ResNet50V2*: Same as in the previous two sections, detailed classification results of Resnet50V2 are presented in this section, Table-II.

ResNet50V2 scores 97% overall accuracy on the four road condition categories. Paved-big, Not-paved-small and Not-paved-big areas are the most well classified categories by the model. Table-I.

Table II: Resnet50V2 classification results

	precision	recall	f1-score	Support
Paved-big	1.00	0.96	0.98	1500
Paved-small	0.94	0.97	0.95	1500
Not-paved-big	0.96	0.96	0.96	3000
Not-paved-small	0.96	0.98	0.97	6000
Accuracy				0.97

3) *EfficientNet-B3*: Below we present detailed classification results of EfficientNet-B3, Table-III.

Table III: EfficientNet-B3 classification results

	precision	recall	f1-score	Support
Paved-big	1	1	1	1500
Paved-small	1	0.97	0.98	1500
Not-paved-big	0.96	1	0.98	3000
Not-paved-small	1	0.98	0.99	6000
Accuracy				0.99

The results in Table-III show that EfficientNet-B3 scores 99% overall accuracy on the four road condition categories. Paved-big and Paved-small are the most well classified categories by the model.

Table-II shows that EfficientNet-B3 outperforms the rest of deep learning models by keeping a high accuracy across categories, especially for paved-big, not-paved-big categories.

### B. Mobility Analytics

Using **fully anonymized** CDR data and the approach discussed in section II-F, we observe the number of trips from origin district, where trips originated from, to destination district, where trip ended. Below we visualize the average percentage volume of origin to destination daily trips per pair of districts, Figure 9. A total number of 1,333,243 trips was estimated for all pairs of districts.

The results show that capital city districts, Gasabo, Kicukiro, Nyarugenge, have more trips in and out, compared to the rest of districts in the country. In Table IV, we combine

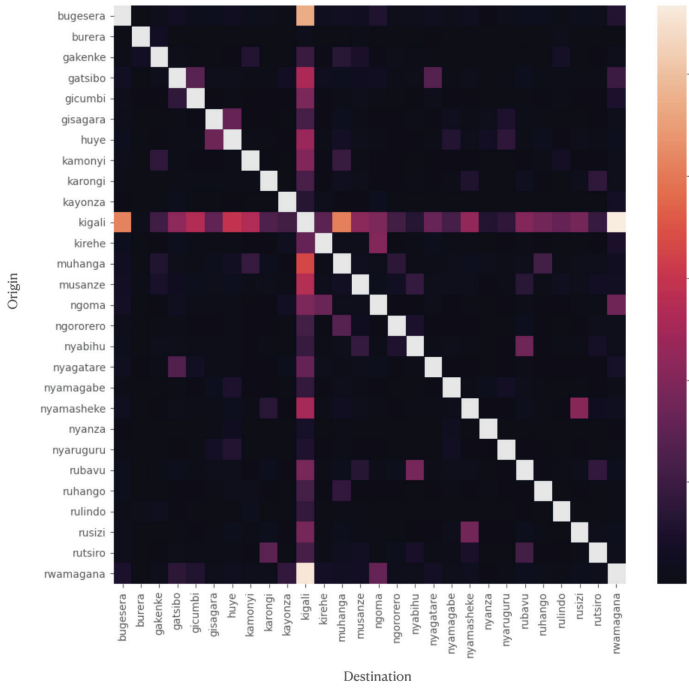


Figure 9: Average origin to destination daily % of total journeys per district

capital city districts into one city, Kigali city, and rank the top 10 districts per number of incoming trips to have the overall picture of famous travel location preference in the country.

Table IV: Top 10 Travel Locations in Rwanda, March 2020

Destination	Estimated Incoming Trips (% of total)
Kigali (capital city)	34.4
nyagatare	4.7
rwamagana	4.6
rubavu	4.4
musanze	4.0
kayonza	3.9
rusizi	3.5
kamonyi	3.3
gatsibo	3.2
bugesera	3.2

### C. Route Infrastructure Prioritization

Considering neighboring districts, we use Google Maps APIs to extract route location, distance and time to travel from one district to its neighboring districts. Then we apply the algorithm discussed in the methodology section II-G to localize not-well connected inter-city routes for neighboring districts, Figure 10.

Results show that districts which are not well connected between them are: Bugesera-Kamonyi, Nyanza-Gisagara, Nyamagabe-Nyamasheke, Nyabihu-Rutsiro, Nyagatare-Gicumbi. Table V.

Using inter-city suggested roads retrieved for neighboring districts, we assess not-paved roads, using the best performing model in section II-E, EfficientNet-B3. Neighboring districts connected with a not-paved road are automatically included

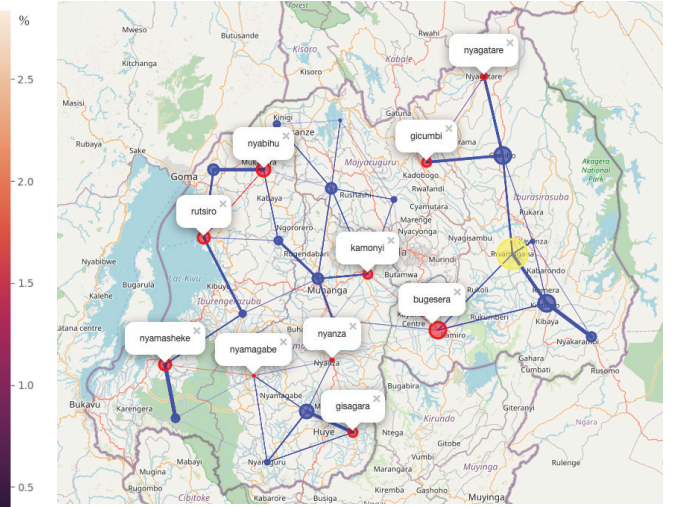


Figure 10: Not-well connected Districts Routes by Travel Time

Table V: Not-well connected districts routes by travel time

District A	District B	Estimated Monthly Trips (group %)
Nyabihu	Rutsiro	55.1
Bugesera	Kamonyi	18.8
Nyagatare	Gicumbi	13.6
Nyamagabe	Nyamasheke	12.5
Nyanza	Gisagara	11.7

among districts with not-well connected routes. See results in Figure 11.



Figure 11: Not-well connected Districts Routes Including Not-paved Roads

By doing this, the following districts are added to neighboring districts with not-well connected routes: Nyabihu-Gakenke, Ngororero-Gakenke, Ngororero-Karongi, Nyamagabe-Karongi, Bugesera-Nyanza, Ruhango-Karongi, Rusizi-Nyaruguru. Table VI.

Given an identified pair of districts with bad road conditions we compute RPS as discussed in section II-H to assess which route could be given high priority for construction, upgrade

Table VI: Not-well connected districts routes due to unpaved roads

District A	District B	Estimated Monthly Trips (group %)
Nyabihu	Gakenke	45.2
Ngororero	Gakenke	13.8
Ngororero	Karongi	13.0
Bugesera	Nyanza	12.5
Ruhango	Karongi	6.4
Nyamagabe	Karongi	6.2
Rusizi	Nyaruguru	2.8

or redesign. Table VII. Given that Rwanda is a small country, we set the distance disadvantage coefficient from famous travel location cities to 0.

Table VII: Route infrastructure prioritization score ( $b_2 = 0$ )

District A	District B	RPS
Nyagatare	Gicumbi	0.13210008
Bugesera	Kamonyi	0.12905084
Bugesera	Nyanza	0.11334041
Nyabihu	Rutsiro	0.10820377
Nyabihu	Gakenke	0.09263452
Rusizi	Nyaruguru	0.08622971
Nyamagabe	Nyamasheke	0.08571947
Ruhango	Karongi	0.08380273
Nyamagabe	Karongi	0.083089
Ngororero	Karongi	0.08066083
Ngororero	Gakenke	0.07186639
Nyanza	Gisagara	0.06658903

Results show that top five routes which would be prioritized using this approach are, Nyagatare-Gicumbi, Bugesera-Kamonyi, Bugesera-Nyanza, Nyabihu-Rutsiro, Nyabihu-Gakenke. With  $b_2 = 1$ , for our case, only the scores would be affected but not the ranks as all of the roads selected using Algorithm-1, turned out to be also unpaved, Table VIII.

Table VIII: Route infrastructure prioritization score ( $b_2 = 1$ )

District A	District B	RPS
Nyagatare	Gicumbi	1.13210008
Bugesera	Kamonyi	1.12905084
Bugesera	Nyanza	1.11334041
Nyabihu	Rutsiro	1.10820377
Nyabihu	Gakenke	1.09263452
Rusizi	Nyaruguru	1.08622971
Nyamagabe	Nyamasheke	1.08571947
Ruhango	Karongi	1.08380273
Nyamagabe	Karongi	1.083089
Ngororero	Karongi	1.08066083
Ngororero	Gakenke	1.07186639
Nyanza	Gisagara	1.06658903

#### IV. CONCLUSION

Assessing road conditions and prioritizing their construction in East Africa, has long relied on the use of structured data and traditional survey approaches for collecting information such as questionnaires, interviews and field visits. These types of surveys are often slow, costly and prone to human error. Using Google Maps APIs we retrieve connecting routes location between neighboring districts to identify not-well connected routes given travel time and road conditions. By using Deep Learning methods and satellite imagery, in this research study we are able to assess with 99% accuracy four different types

of road conditions. Namely, Paved-big, Paved-small, Not-paved Big, Not-paved-small. The best performing model for assessing these road conditions in terms of an accuracy of 99% was EfficientNet-B3.

Given districts with bad road conditions and not-well connected routes, we use extracted mobility patterns from CDR data to provide a data oriented method on how efficiently their construction and maintenance could be prioritized. Among not-well connected routes the following pair of districts rank higher in terms of construction and maintenance prioritization: Nyagatare-Gicumbi, Bugesera-Kamonyi, Bugesera-Nyanza, Nyabihu-Rutsiro, Nyabihu-Gakenke.

The results show that given a specific country in East Africa, one could assess with confidence inter-city road conditions suggested by Google Maps API and rank them for infrastructure prioritization.

On the other hand this highlights that Google Maps API does not always suggest the roads with better conditions in the context of Africa (paved vs not-paved), rather the ones with less travel time. With the methodology provided in this research it's possible to find a better traveling route network by avoiding roads with bad conditions given classification results.

Classification results in this research express that the deep learning models used were able to learn different patterns independently by observing the ranking accuracy per category. Additional studies of blended optimal deep model architectures for classification of more variety of road conditions would be interesting to explore.

#### REFERENCES

- [1] Rwanda Transport Development Agency. *Map of Classified National and District Roads Class 1 Network 2020*. <https://www.rtda.gov.rw>. [Online; accessed March-15-2021]. 2020.
- [2] Sruthy Agnisarman et al. "A survey of automation-enabled human-in-the-loop systems for infrastructure visual inspection". In: *Automation in Construction* 97 (2019), pp. 52–76.
- [3] Facundo Alvaredo et al. *World inequality report 2018*. Belknap Press, 2018.
- [4] Kathryn A Arnold, Alize Le Roux, and Sibusisiwe A Makhanya. "Implementing a GIS based methodology for determining highly vulnerable rural access roads to a changing climate in Ethiopia". In: (2018).
- [5] Rodrigo F Berriel et al. "Deep learning-based large-scale automatic satellite crosswalk classification". In: *IEEE Geoscience and Remote Sensing Letters* 14.9 (2017), pp. 1513–1517.
- [6] Benjamin Bischke et al. "Multi-task learning for segmentation of building footprints with deep neural networks". In: *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2019, pp. 1480–1484.
- [7] Kenneth K Boyer, John R Olson, and Eric C Jackson. "Electronic surveys: Advantages and disadvantages over traditional print surveys". In: *Decision Line* 32.4 (2001), pp. 4–7.

- [8] Gabriel Cadamuro, Aggrey Muhebwa, and Jay Taneja. "Assigning a grade: Accurate measurement of road quality using satellite imagery". In: *arXiv preprint arXiv:1812.01699* (2018).
- [9] Xavier Espinet and Julie Rozenberg. "Prioritization of climate change adaptation interventions in a road network combining spatial socio-economic data, network criticality analysis, and flood risk assessments". In: *Transportation Research Record* 2672.2 (2018), pp. 44–53.
- [10] Mihály Fazekas and Bence Tóth. "The extent and cost of corruption in transport infrastructure. New evidence from Europe". In: *Transportation research part A: policy and practice* 113 (2018), pp. 35–54.
- [11] Stephen Gibbons et al. "New road infrastructure: the effects on firms". In: *Journal of Urban Economics* 110 (2019), pp. 35–50.
- [12] Yanming Guo et al. "Deep learning for visual understanding: A review". In: *Neurocomputing* 187 (2016), pp. 27–48.
- [13] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [14] Kaiming He et al. "Identity mappings in deep residual networks". In: *European conference on computer vision*. Springer. 2016, pp. 630–645.
- [15] Yihui He et al. "Amc: Automl for model compression and acceleration on mobile devices". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 784–800.
- [16] Ilaria Henke et al. "Decision-making in the transport sector: A sustainable evaluation method for road infrastructure". In: *Sustainability* 12.3 (2020), p. 764.
- [17] Mohamad Ali Khalil et al. "Multi-Temporal satellite imagery for infrastructure growth assessment of Dubai City, UAE". In: *MATEC Web of Conferences*. Vol. 120. EDP Sciences. 2017, p. 09006.
- [18] Muhammad Jaleed Khan et al. "Automatic target detection in satellite images using deep learning". In: *J. Space Technol* 7.1 (2017), pp. 44–49.
- [19] Laurens van der Maaten and Geoffrey Hinton. "Visualizing data using t-SNE". In: *Journal of machine learning research* 9.Nov (2008), pp. 2579–2605.
- [20] Darwin Marcelo et al. "Prioritizing water supply infrastructure investments in Sri Lanka: an application of the world bank infrastructure prioritization framework". In: *World Bank Policy Research Working Paper* 8331 (2018).
- [21] Bongeka L Mbutuma, M Mostafa Hassan, and DK Das. "Impact of Climate Change on Traffic Controls in Developing Countries". In: *International Congress and Exhibition "Sustainable Civil Infrastructures: Innovative Infrastructure Geotechnology"*. Springer. 2018, pp. 46–57.
- [22] Zeynu Mohammed, Mesfin Mekonnin, and Abera Jaleta. "Factors affecting logistics performance emphasis on transport function: a study on east africa bottling sc jimma branch". PhD thesis. 2020.
- [23] Mama Nsangou Mouchili, John William Atwood, and Shadi Aljawarneh. "Call data record based big data analytics for smart cities". In: *Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems*. 2019, pp. 1–7.
- [24] Esdras Ngezahayo, Gurmeh S Ghataora, and Michael PN Burrow. "Factors Affecting Erosion in Unpaved Roads". In: *Proc. of the 4th Intern. Conf. on Geotechnical Research and Eng.* 2019, pp. 7–9.
- [25] Barak Oshri et al. "Infrastructure quality assessment in africa using satellite imagery and deep learning". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2018, pp. 616–625.
- [26] O Petrucci and AA Pasqua. "Damaging events along roads during bad weather periods: a case study in Calabria (Italy)". In: *Natural Hazards and Earth System Sciences* 12.2 (2012), pp. 365–378.
- [27] Ming Shan et al. "Investigating the underlying factors of corruption in the public construction sector: Evidence from China". In: *Science and engineering ethics* 23.6 (2017), pp. 1643–1666.
- [28] Marcin Staniek. "Road pavement condition as a determinant of travelling comfort". In: *Intelligent Transport Systems and Travel Behaviour*. Springer, 2017, pp. 99–107.
- [29] Christian Szegedy et al. "Rethinking the inception architecture for computer vision". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826.
- [30] Haifa Tamiminia et al. "Google Earth Engine for geobig data applications: A meta-analysis and systematic review". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 164 (2020), pp. 152–170.
- [31] Chuanqi Tan et al. "A survey on deep transfer learning". In: *International conference on artificial neural networks*. Springer. 2018, pp. 270–279.
- [32] Mingxing Tan and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks". In: *International Conference on Machine Learning*. PMLR. 2019, pp. 6105–6114.
- [33] Reza Tolouei, Stefanos Psarras, and Rawle Prince. "Origin-destination trip matrix development: Conventional methods versus mobile phone data". In: *Transportation research procedia* 26 (2017), pp. 39–52.
- [34] Grigorios Tsagkatakis et al. "Survey of deep-learning approaches for remote sensing observation enhancement". In: *Sensors* 19.18 (2019), p. 3929.
- [35] Adam Van Etten. "You only look twice: Rapid multi-scale object detection in satellite imagery". In: *arXiv preprint arXiv:1805.09512* (2018).
- [36] Athanasios Voulodimos et al. "Deep learning for computer vision: A brief review". In: *Computational intelligence and neuroscience* 2018 (2018).
- [37] Li Zhu et al. "Big data analytics in intelligent transportation systems: A survey". In: *IEEE Transactions*

on *Intelligent Transportation Systems* 20.1 (2018), pp. 383–398.



**Davy K. Uwizera** is a Masters of Science in Information Technology graduate from Carnegie Mellon University, class of 2017, currently pursuing a dual PhD in data science at Africa Center of Excellence in Data Science and a PhD in Internet of Things at Africa Center of Excellence in Internet of Things. His current research interests include unstructured data mining, deep learning, machine learning edge and big data.



**Charles Ruranga** Received the B.Sc. degree in economics from National University of Rwanda, Rwanda, 2001, M.Sc. degree in economics, University of Montreal, Canada, 2007. He received his PhD in statistics from Makerere University, Uganda in 2017. His research interests include, but are not limited to, econometric spatial analysis, time series and multivariate signal processing. He is currently the director of the African Center of Excellence in Data Science (ACE-DS).



**Patrick McSharry** (Senior Member of the IEEE) He received the B.A. degree in theoretical physics in 1993 and the M.Sc. degree in electronic and electrical engineering in 1995 from Trinity College Dublin, Dublin, Ireland. He received the D.Phil. degree in mathematics from the university of Oxford, U.K., in 1999. He is currently visiting professor at Carnegie Mellon University Africa and a member of the Kigali Collaborative Research Center. Having served 22 years at Oxford University, McSharry remains affiliated with the Oxford Man Institute of

Quantitative Finance, the Mathematical Institute, the Department of Computer Science, and the Oxford Internet Institute. He is a Fellow of the Royal Statistical Society, Senior Member of the IEEE and Senior Academic Member of the Willis Research Network.