

Thegeya, Aaron et al.

Working Paper

Application of machine learning algorithms on satellite imagery for road quality monitoring: An alternative approach to road quality surveys

ADB Economics Working Paper Series, No. 675

Provided in Cooperation with:

Asian Development Bank (ADB), Manila

Suggested Citation: Thegeya, Aaron et al. (2022) : Application of machine learning algorithms on satellite imagery for road quality monitoring: An alternative approach to road quality surveys, ADB Economics Working Paper Series, No. 675, Asian Development Bank (ADB), Manila, <https://doi.org/10.22617/WPS220587-2>

This Version is available at:

<https://hdl.handle.net/10419/298121>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/3.0/igo/>

APPLICATION OF MACHINE LEARNING ALGORITHMS ON SATELLITE IMAGERY FOR ROAD QUALITY MONITORING

AN ALTERNATIVE APPROACH TO ROAD QUALITY SURVEYS

*Aaron Thegeya, Thomas Mitterling, Arturo Martinez Jr, Joseph Albert Niño Bulan,
Ron Lester Durante, and Jayzon Mag-atas*

NO. 675

December 2022

**ADB ECONOMICS
WORKING PAPER SERIES**

ADB Economics Working Paper Series

Application of Machine Learning Algorithms on Satellite Imagery for Road Quality Monitoring: An Alternative Approach to Road Quality Surveys

Aaron Thegeya, Thomas Mitterling,
Arturo Martinez Jr, Joseph Albert Niño Bulan,
Ron Lester Durante, and Jayzon Mag-atas

No. 675 | December 2022

The *ADB Economics Working Paper Series* presents research in progress to elicit comments and encourage debate on development issues in Asia and the Pacific. The views expressed are those of the authors and do not necessarily reflect the views and policies of ADB or its Board of Governors or the governments they represent.

Aaron Thegeya (athegeya@gmail.com) and Thomas Mitterling (thomas.mitterling@worlddata.io) are data scientists at the World Data Lab. Arturo Martinez Jr. is a statistician (amartinezjr@adb.org), Joseph Bulan (jbulan@adb.org) is an associate statistics analyst, and Ron Lester Durante (rdurante1.consultant@adb.org) and Jayzon Mag-atas (jmagatas.consultant@adb.org) are consultants at the Economic Research and Regional Cooperation Department, Asian Development Bank.



Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO)

© 2022 Asian Development Bank
6 ADB Avenue, Mandaluyong City, 1550 Metro Manila, Philippines
Tel +63 2 8632 4444; Fax +63 2 8636 2444
www.adb.org

Some rights reserved. Published in 2022.

ISSN 2313-6537 (print), 2313-6545 (electronic)
Publication Stock No. WPS220587-2
DOI: <http://dx.doi.org/10.22617/WPS220587-2>

The views expressed in this publication are those of the authors and do not necessarily reflect the views and policies of the Asian Development Bank (ADB) or its Board of Governors or the governments they represent.

ADB does not guarantee the accuracy of the data included in this publication and accepts no responsibility for any consequence of their use. The mention of specific companies or products of manufacturers does not imply that they are endorsed or recommended by ADB in preference to others of a similar nature that are not mentioned.

By making any designation of or reference to a particular territory or geographic area, or by using the term “country” in this document, ADB does not intend to make any judgments as to the legal or other status of any territory or area.

This work is available under the Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO) <https://creativecommons.org/licenses/by/3.0/igo/>. By using the content of this publication, you agree to be bound by the terms of this license. For attribution, translations, adaptations, and permissions, please read the provisions and terms of use at <https://www.adb.org/terms-use#openaccess>.

This CC license does not apply to non-ADB copyright materials in this publication. If the material is attributed to another source, please contact the copyright owner or publisher of that source for permission to reproduce it. ADB cannot be held liable for any claims that arise as a result of your use of the material.

Please contact pubsmarketing@adb.org if you have questions or comments with respect to content, or if you wish to obtain copyright permission for your intended use that does not fall within these terms, or for permission to use the ADB logo.

Corrigenda to ADB publications may be found at <http://www.adb.org/publications/corrigenda>.

Note:

In this publication, “\$” refers to United States dollars.

The ADB Economics Working Paper Series presents data, information, and/or findings from ongoing research and studies to encourage exchange of ideas and to elicit comment and feedback about development issues in Asia and the Pacific. Since papers in this series are intended for quick and easy dissemination, the content may or may not be fully edited and may later be modified for final publication.

ABSTRACT

Roads are vital to support the transportation of people, goods, and services, among others. To yield their optimal socioeconomic impact, proper maintenance of existing roads is required; however, this is typically underfunded. Since detecting road quality is both labor and capital intensive, information on it is usually scarce, especially in resource-constrained countries. Accordingly, the study examines the feasibility of using satellite imagery and artificial intelligence to develop an efficient and cost-effective way to determine and predict the condition of roads. With this goal, a preliminary algorithm was created and validated using medium-resolution satellite imagery and existing road roughness data from the Philippines. After analysis, it was determined that the algorithm had an accuracy rate up to 75% and can be used for the preliminary identification of poor to bad roads. This provides an alternative for compiling road quality data, especially for areas where conventional methods can be difficult to implement. Nonetheless, additional technical enhancements need to be explored to further increase the algorithm's prediction accuracy and enhance its robustness.

Keywords: road quality, road maintenance, Sustainable Development Goals, remote sensing, deep learning

JEL codes: O18, R42

This paper was prepared for the technical assistance project funded by the Japan Fund for Prosperous and Resilient Asia and the Pacific entitled Using Frontier Technology and Big Data Analytics for Smart Infrastructure Facility Planning and Monitoring. The authors would like to thank colleagues from the Philippine Department of Public Works and Highways, who worked closely with the project team; the Philippine Institute for Development Studies; and the participants of the 15th National Convention on Statistics for providing insightful feedback that helped refine the findings of the study. Michael Ogolla, ADB consultants Mildred Addawe, Raymond Realeza Adofina, Kristine Faith Agtarap, Maria Concordia Alfonso, Remedios Baes-Espineda, Nalwino Billones, Madeline Cabauatan, Dennis Dizon, Anna Marie Fernando, Rica Cynthia Maddawin, Marymell Martillan, and Christian Flora Mae Soco, all contributed to works that were used as inputs and references for this report. Iva Sebastian, with the assistance of Ma. Roselia Babalo, Aileen Gatson, and Rose Anne Dumayas provided operational support through the course of the project. Manuscript editing was performed by Jason Beerman.

I. INTRODUCTION

A. Background

Roads are critical conduits that enable the flow of economic activity in every country and are an indispensable component of every country's physical infrastructure. Roads facilitate the movement of goods and services, connect rural areas to urban areas and economic centers, enable easy access to education and health care, and facilitate the free mobility of labor and ideas within and across countries. The importance of access to good quality roads is further highlighted in a development target 9.1 set by the Sustainable Development Goals, which aims to “develop quality, reliable, sustainable and resilient infrastructure, including regional and transborder infrastructure, to support economic development and human well-being, with a focus on affordable and equitable access for all” (United Nations 2015).

Given the socioeconomic impact of roads, governments spend significant portions of their budgets to build and maintain road networks, and other types of infrastructures, in general. A study conducted by the Asian Development Bank (2017) estimates that for developing Asia to maintain its growth momentum, tackle poverty, and respond to climate change, Developing Asia needs to invest \$1.7 trillion per year in infrastructure until 2030. Of this amount, \$8.4 trillion is needed for transport infrastructure.

Since 2011, the Philippines has adopted a policy defining 20-year design life for concrete roads and 10-year design life for asphalt road (Department of Public Works and Highways 2011). Nonetheless, roads should not reach the end of their design life to retain their serviceability and avoid the need for reconstruction or replacement. For the Philippines, major maintenance activities are ideally conducted by the Department of Public Works and Highways for national roads every 10 years for concrete pavement and every 5 years for asphalt pavement, while the maintenance of local roads is handled by local government units. However, the need for maintenance may vary depending on the external factors experienced by the road such as rainfall, temperature, and traffic volume, among others, which may affect the deterioration of the road. Given these conditions, road maintenance and regular monitoring are key elements in ensuring that transport infrastructures yield optimal socioeconomic impact; however, these elements are typically underfunded (Burningham and Stankevich 2005). On average, countries spend 20%–50% of what they should be spending on maintenance, and this is compounded by the costs of collecting data to identify roads in need of maintenance. Given that detecting the quality of roads is both capital and labor intensive, data on the quality of roads is usually scarce, especially in resource-constrained countries.

The Rural Access Index, an indicator under Sustainable Development Goal 9, measures the proportion of the rural population living within 2 kilometers of an all-season road. This index may provide insights about areas where road infrastructure gaps exist, particularly in rural settings. Compiling the Rural Access Index requires three types of geospatially tagged data: population, road networks, and all-season roads. However, there are several technical challenges when collecting data about all-season roads. First, the concept used can be operationalized in different ways. Some countries assess the all-season status based on visual assessment of road conditions, others on the average speed of vehicles, while others on roughness, making intercountry comparisons challenging (Workman and McPherson 2021). Second, collecting the data itself is also challenging. For instance, conventional methods of detecting road quality include the measurement and collection of road survey instrument data. The international roughness index (IRI) is a common metric that is used in several countries to measure road quality and is

calculated using a mathematical simulation of a single car wheel traveling along a given road predetermined speed. Road quality surveys are expensive to undertake and are therefore rarely used, particularly in low-income countries. Using smartphones to capture images of roads offers promise, although this solution also presents its own challenges. Such an approach is likely to be limited to only urban areas with high bandwidth and internet connectivity.

Considering the constraint of resources in middle- and low-income areas, this paper proposes an alternative approach to detecting road quality that leverages artificial intelligence and advances in satellite imagery technologies. The proliferation of institutions with access to satellite images has made it possible to capture high-resolution images of up to 30–50 centimeters in any region in the world. Moreover, developments in artificial intelligence, particularly transfer learning¹ and convolutional neural networks (CNNs)², provide a method to measure road quality while leveraging large datasets. Our approach utilizes data from Google Earth, which are at a lower resolution of 10 meters per pixel, but also leverages other satellite data that are likely to be correlated with road quality, including temperature, precipitation, terrain, and population density.

Using the Philippines presents an interesting feasibility study for this proposed alternative approach of compiling data on road conditions. To address the infrastructure deficit and for the continuous growth of the country, the Government of the Philippines has an initiative that prioritizes allotting investments on infrastructure development. It covers various projects that target specific aspects of the Philippines' transportation infrastructure, such as the construction of new (or the improvement of existing) roads, bridges, ports, transport terminals, flood control infrastructure, and communication facilities. As access to roads expands under such projects, it is also important to explore cost-effective approaches of monitoring the condition of these infrastructures over time.

The approach uses three types of neural networks to extract features of road quality from satellite images of roads. These include CNNs trained over road image data, neural networks trained on tabular data, and a combined neural network trained on both visual and tabular data. Covariates likely to affect road quality such as gradient level, precipitation, population density, and temperature variation are collected from satellite sources, and are supplemented with information collected from shapefiles³ on road classification and pavement type. Thereafter, a combined CNN that takes concatenated image and tabular data is used to classify the roads.

This approach has the following advantages over conventional survey methods of measuring road quality: (i) it is cheaper and therefore allows for more frequent collection of data on road quality, (ii) it is less labor intensive and therefore can be used by countries that do not have skilled personnel in road quality monitoring, and (iii) it allows dynamic measuring of road quality over time.

¹ Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task.

² A Convolutional Neural Network (CNN) is a deep learning algorithm that can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and be able to differentiate one from the other.

³ A shapefile is a simple, nontopological format for storing the geometric location and attribute information of geographic features.

B. Literature Review

Literature on the identification of road quality has seen significant advancement with the advent of machine learning algorithms (MLAs) and, in particular, the use of neural networks trained on large datasets. A large literature exists on methods for automated processing of large datasets of instrument data, on road segmentation using unsupervised learning of image data, and (most recently) on the determination of the quality of roads.

Cadamuro, Muhebwa, and Taneja (2018) built a CNN to predict road quality in Kenya, using IRI data as a proxy for road quality. Their study was limited to 1,150 kilometers of data, where they predicted five classes of roads based on IRI data, as well as a binary classification model. They predicted outcomes using three CNN variants: AlexNet,⁴ SqueezeNet,⁵ and Visual Geometry Group (VGG).⁶ Their models achieved an accuracy of at least 87% on the binary classification model, but performance fell on the five-class variant of the problem to 64%. The homogeneity of any given section of road has positive correlation with the ability to predict road quality, suggesting that a sequential approach may yield better results. Our study mirrors this approach, using publicly available data from the Google Earth Engine platform.

Mnih and Hinton (2010) trained one of the first neural networks to detect roads from high-resolution aerial imagery obtained from road maps. The road map data specified the centerline of each road, and the authors approximated pixels within the width of the road. Their neural network was pretrained using a procedure that made use of restricted Boltzmann machines, and further trained using labeled road map data. Their results were better than any preexisting published research based on automatic road detection.

Oshri et al. (2018) analyzed several types of infrastructure throughout Africa, including roadways, to predict infrastructure accessibility and quality. They combined infrastructure assessments from Afrobarometer data with satellite imagery. Using a residual neural network (ResNet) with transfer learning, they were able to correctly classify roadway quality 70.5% of the time. Their predictive accuracy was robust to separately train the model on urban and rural areas as well as across countries, although the authors noted concerns about potential overfitting.

Gerke and Heipke (2008) utilized the sequential nature of roadway networks to improve the extraction of roadway imagery from satellite imagery. They first applied standard computer vision techniques to build a baseline map of roadways, and then used the resulting network structure to fill in gaps and improve the accuracy of image extraction. Their paper suggests that a network or sequential based approach may be useful for predicting road quality, by utilizing the contiguity of road segments.

Fiorentini et al. (2021) compared predictions of vertical displacement measures of road quality by MLAs with road roughness surveys in Italy. The aim of the study was to examine the robustness of MLA predictions of road quality to establish whether they can replace expensive and time-

⁴ AlexNet has eight layers with learnable parameters. The model consists of five layers with a combination of max pooling followed by three fully connected layers (Krizhevsky, Sutskever and Hinton 2012).

⁵ SqueezeNet is a CNN architecture designed to have an equivalent accuracy to AlexNet with lesser parameters. As a result, it requires less communication across servers during distributed training and less bandwidth to export a new model from the cloud to an autonomous car. It is also more feasible to deploy on field programmable gate arrays and other hardware with limited memory (Iandola et al. 2016).

⁶ VGG is a deep CNN architecture that aimed to improve the accuracy in the large-scale image recognition setting by pushing the depth up to 19 weight layers (Simonyan and Zisserman 2014).

consuming profilometric road surveys. Further, the study applied a persistent scatter interferometric algorithm to obtain data from satellite images and compared these to IRI values from a laser profiler. The study found that significant correlations between conventional IRI values and MLA predictions were associated with exogenous factors, while weak correlations were attributed to endogenous (local) factors such as traffic loading. The authors recommend calibrating MLAs with endogenous (local) road conditions to improve prediction accuracy.

Leduc and Assaf (2022) used 298 road images captured using GoPro cameras at an angle of 60 degrees to identify and classify road deformations. The authors used an automated machine learning model and CNN to identify and classify road deformations. The study showed that the automated machine learning ranges from 0.50 to 0.92. Visual data used to validate the machine learning (ML) results showed 71% of true positives.

Marcelino, Lurdes Antunes, and Fortunato (2018) evaluated road pavement conditions on a 157-kilometer motorway in Portugal. The study used continuous data such as IRI, mean profile depth, and cracking area. Further, the study applied a regularized regression with a lasso algorithm to predict an indicator for road quality. The results showed models from the MLA had lower mean squared errors than conventional methods of examining pavement conditions. The study concluded that machine learning techniques are a promising solution for predicting road quality relative to conventional road evaluation methods.

Bashar and Torres-Machi (2021) applied artificial neural network (ANN), random forest, and support vector machine (SVM) algorithms to predict IRI scores, using data from a number of selected studies. Random forests showed the best results in predicting IRI scores with an overall performance of 0.995, while ANN had consistently accurate results over a significant number of studies. SVM algorithms showed completely different results to ANN and random forest models.

Eisenbach et al. (2019) analyzed nearly 2,500 images of pavement taken from German highways to predict the distress level of the pavements. They trained three neural networks: VGG, ResNets, and SVM on a binary decision problem (distress versus no distress). Both ResNets and VGG networks outperformed classical machine learning models like SVMs in prediction accuracy. In conclusion, ResNet was found to be the most suitable for detecting pavement distress since it uses less computational power.

Abdelaziz et al. (2020) utilized long-term pavement performance data from the Government of the United States to predict IRI classification on highway data from the United States. They found that a standard CNN outperforms a linear regression classification model, with an R-squared of 0.75 compared to 0.57.

The literature highlights three challenges faced in conducting research using image data. First is the level of resolution of satellite imagery, whereby poor resolution data will not show road defects. Second, there may be a temporal mismatch between survey data used for training purposes and satellite imagery, which could be highly problematic given that road quality can change drastically over a short period because of weather or construction. Third, road quality data may be sequential in nature, restricting the assignment of data into train and test datasets. This requires the division of data into longer sections before assignment to each dataset. Existing literature suggests that using CNNs to predict road quality has enjoyed some success. Neural networks have been more successful and have generally outperformed other classification methods. A key factor is the quantity and quality of data, which affects model performance.

II. Materials and Methods

A. Data

The international roughness index data, and remotely sensed and modeled data are utilized in the model:

1. International Roughness Index Data

The IRI, introduced by the World Bank in the 1980s, is commonly used (Bennett, De Solminihaç, and Chamorro 2006) in several countries to measure road quality. The IRI measures how an idealized vehicle responds to a given road's profile and is calculated using a mathematical simulation of a single car wheel traveling along the road profile at a predetermined speed, usually 80 kilometers per hour. The IRI expresses a ratio of the accumulated suspension motion of a vehicle, divided by the distance traveled during the test (Gillespie, Paterson, and Sayers 1986). The data may be measured through a variety of methods, such as the vertical distance that a laser mounted on a profiler van jumps as the van moves along a road. The IRI is standardized and measured in one of three units: meters per kilometer, inches per mile, and millimeters per meter. There are different scales of measuring the IRI such as response type road roughness meters and the present serviceability rating. A standard scale that contains IRI values from 0 to 5 is commonly used to measure IRI road values.

IRI data are obtained for a collection of roads in the Philippines, covering a total of 15 regions within 3 major island groups: Luzon, Visayas, and Mindanao. Road identifier data contain road names, as well as geographical identifiers of the island, region, province, and congressional district where the road is located. IRI data were collected as of 2019. Road quality data are collected for road segments that are roughly 100 meters long, for a total of 124,462 observations.⁷ Road segments are also grouped into a total of 1,046 road sections. Road quality is measured with an average IRI reading for each 100-meter segment, which is also translated into an IRI rating. In total, there are four IRI ratings: bad, poor, fair, and good; the corresponding IRI range of values is shown in Table 1 (Department of Public Works and Highways 2021).

Table 1: International Roughness Index Rating

IRI Rating	IRI Value Range
Bad	>7
Poor	>5–7
Fair	>3–5
Good	1–3

IRI = international roughness index.

Source: Government of the Philippines, Department of Public Works and Highways.

Road segments are numbered sequentially, therefore allowing the identification of adjacent segments. In addition, all roads are classified as either primary or secondary. In addition to the IRI data, other information is provided within additional shapefiles, which are merged with the IRI data. The data includes road carriage number of lanes, surface type, pavement type, width, flow type, and type of road shoulder.

⁷ Road segments in the dataset range from 1 to 107 meters, with most segments ranging from 99 to 101 meters. The last segment measured for any given road is most likely to have a length less than 100 meters.

2. Remotely Sensed and Modeled Data

The remotely sensed data source considered for the analysis is the Copernicus program, headed by the European Space Agency. Copernicus consists of the Sentinel satellites, which are a constellation of satellites that take weather radar images, medium-resolution optical images, ocean data, and land data to monitor the environment, climate, and air quality. The study used the medium-resolution imageries taken in 2019 from the Sentinel-2 (Copernicus Sentinel Data 2021) multispectral satellite imaging mission with a global 5-day revisit frequency. The Sentinel-2 Multispectral Instrument samples 13 spectral bands: visible and near-infrared at a resolution of up to 10 meters, red-edge, and shortwave infrared at a resolution of 20 meters, and atmospheric at a resolution of 60 meters.

First, remotely sensed data were collected for all relevant road segments using publicly accessible Google Earth application programming interfaces (APIs) (Figure 1). To match the remotely sensed data to the IRI data, latitude, and longitude coordinates of the road's bounding box are obtained.⁸ The bounding box is then used to isolate the corresponding Sentinel-2 imagery from Google Earth within a 90-day window centered on the date the IRI was surveyed. The collection of images within this window that has a maximum cloud cover of 6% is then downloaded, and identified cloudy pixels are masked. The median value for each pixel is then selected across all images in the image collection, thereby producing a single image. The image is then cropped to include only the pixels within the bounding box, which are selected with a resolution of 10 meters per pixel, which is Sentinel-2's resolution.⁹ A 90-day window is selected to ensure that a significant number of images is available for each road segment, given the cloud cover parameters, as well as the frequency with which the satellites visit each location.

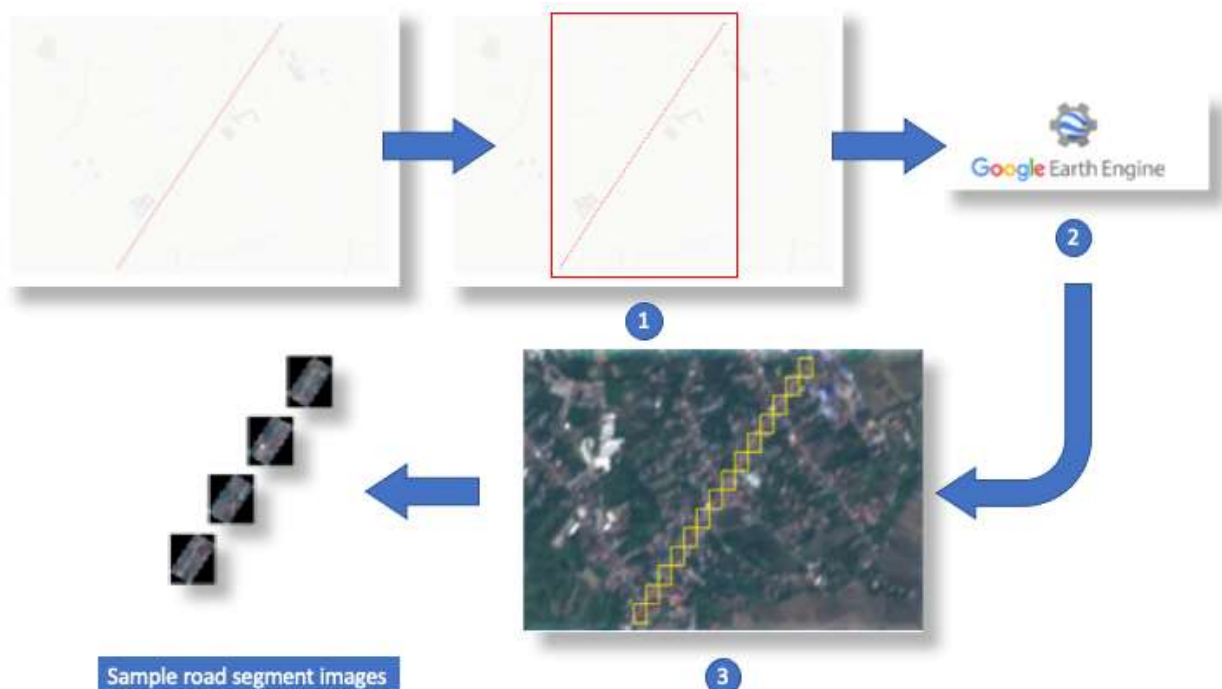
Lower image resolution means that image contrast is restricted to a small number of pixels. On the other hand, satellite data utilized in Cadamuro, Muhebwa, and Taneja (2018) has a typical resolution of 0.5 meters per pixel, which is 20 times higher. A 100-meter segment is therefore covered by fewer than 10 pixels with publicly available data, compared to 200 pixels from a high-resolution data source. A rule of thumb is that higher resolution is generally better for analysis, although the authors are not aware of any literature concerning the impact of resolution on the accuracy of road classification. It is instructive to note that from a visual perspective, individual defects within a road, such as surface cracks or potholes, are not visible to the naked eye in both the publicly available and high-resolution satellite data. However, high-resolution data clearly allows identification of the road type. Sabottke and Spieler (2020) find that for binary classification problems, resolutions above 224×224 pixels only provide a modest additional return regarding classification accuracy.

Figure 1 summarizes the process of downloading satellite imagery for each road segment, which is broken down into three steps: (i) coordinates of the road's bounding box were obtained; (ii) the corresponding Sentinel-2 road imagery was downloaded from Google Earth Engine; and (iii) imagery of individual road segments with available IRI data was cropped to the size of its bounding box.

⁸ Image bounding boxes are extracted from polygons within the IRI data shapefiles and imported into Google Earth based on the WGS84 (EPSG:4326) geographic coordinate system. Image data are downloaded from Google Earth in a GeoTIFF format, and thereafter converted into a JPEG format to allow analysis within a CNN.

⁹ As the highest-resolution imagery is downloaded, and there are no other bands with higher resolution than 10 meters per pixel, there is no opportunity to increase resolution through techniques such as panchromatic sharpening.

Figure 1: Satellite Imagery Downloading Flowchart



Source: The flowchart was generated by the authors using basemaps from OpenStreetMap (accessed 04 March 2022) and satellite imagery from Sentinel-2 daytime imagery downloaded using Google Earth Engine (accessed 04 March 2022).

In addition to visual road segment data, the following remotely sensed and modeled data were aggregated within the bounding boxes of every 100-meter road segment to produce the following tabular dataset:

- (i) Average daily temperature at a height of 2 meters, from the European Centre for Medium-Range Weather Forecasts' fifth-generation reanalysis for global climate and weather with a resolution of 0.25 arc degrees.¹⁰
- (ii) Average total precipitation from the European Centre for Medium-Range Weather Forecasts' fifth-generation reanalysis (Copernicus Climate Change Service 2017), with a resolution of 0.25 arc degrees.
- (iii) Land gradient calculated using the Shuttle Radar Topography Mission of the National Aeronautics and Space Administration (NASA) (Farr et al. 2007), with a resolution of 30 meters. Elevation is measured at each corner of the bounding box, and gradient calculated as the difference between the two values.
- (iv) Total population from WorldPop (Gaughan et al. 2013), estimated per 100 meter-by-100 meter grid square. The population within a neighborhood of 1 square kilometer is included to give an indication of the total population within a given area.

The above tabular data supplement visual data for each road segment and are collected as they are likely to influence road quality. Temperature and precipitation are important factors that influence how quickly a road degrades. In addition, land gradient determines land surface flow of

¹⁰ A resolution of 0.24 arc degrees corresponds to about 25 kilometers. For comparison, the average daily land surface temperature from the NASA MODIS satellite (resolution of 1,000 meters) is also used.

water, which impacts road quality. Population within the vicinity of a road segment is an indicator of the frequency of use of a given road segment.

B. Summary Statistics

Data on road quality covers the islands of Luzon, Visayas, and Mindanao (Table 2). Tertiary roads are excluded from the coverage, and the data collected is restricted to only primary and secondary roads. The largest number of road segments is in the Visayas, which accounts for 43% of all road quality data, while Luzon accounts for the smallest number of road segments. The average IRI value of the sample is 4.68, with IRI values across all islands ranging from 1.00 to 17.03. Roads in Mindanao have the highest average IRI value while those in Luzon have the lowest.

Table 2: Road Segment Summary Statistics

Island	No. of Road Segments	Average IRI	Minimum IRI	Maximum IRI
Luzon	28,194	4.10	1.03	17.03
Visayas	53,808	4.75	1.03	14.75
Mindanao	42,460	4.97	1.00	15.17
Total	124,462	4.68	1.00	17.03

IRI = international roughness index.

Source: Government of the Philippines, Department of Public Works and Highways.

Of the roads covered, 62% are rated either poor or fair, while 25% are rated good. In total, 13% of roads are rated bad (Table 3). The highest concentration of roads rated good is in Region IV-A in Luzon, where 57% of roads are rated good. Also, more than half of the roads in Region V in Luzon are rated good. The highest concentration of lower quality roads is also within Luzon, within the Cordillera Administrative Region, where 25% of its roads are rated bad and an additional 46% are rated poor. Region XII and Region XI within Mindanao Island, as well as the National Capital Region, are ranked with a high relative share of roads classified as bad.

Table 3: Regional Summary Statistics
(%)

Location	Average Rating			
	Bad	Poor	Fair	Good
Luzon				
Cordillera Admin Region	25.3	46.4	16.9	11.5
National Capital Region	17.3	27.3	33.3	22.0
Region I	3.8	17.1	39.8	39.3
Region II	9.9	28.8	39.3	22.0
Region III	4.5	16.0	33.2	46.4
Region IV-A	5.7	8.6	29.0	56.7
Region V	4.8	10.9	33.6	50.8
Visayas				
Region VI	14.6	37.6	29.8	18.0
Region VII	8.9	30.3	39.7	21.1
Region VIII	12.8	30.1	30.2	26.8
Mindanao				
Region IX	14.4	35.9	35.4	14.3

Region X	8.8	29.5	28.7	33.0
Region XI	17.1	41.9	24.7	16.4
Region XII	21.8	32.7	22.0	23.4
Region XIII	11.4	26.0	29.9	32.7
Total	12.8	30.8	31.4	25.1

Source: Government of the Philippines, Department of Public Works and Highways.

III. Results

A. Training and Results

Analysis is carried out using three neural network architectures: (i) a baseline CNN is trained on image data following Cadamuro, Muhebwa, and Taneja (2018); (ii) a two-layer neural network is trained on tabular data; and (iii) a combined model is used that includes both visual and tabular data.¹¹ Across the different models, a cross-entropy loss function is implemented to show deviations of model predictions from actual values; this gives a high penalty for incorrect predictions made with a high level of confidence and mean square error.¹²

1. Convolutional Neural Network Model Results

A CNN is a multilayered feed-forward network with a convoluted structure; it generally consists of an input layer, multiple convolution layers, pooling layers, a fully connected layer, and an output layer. The baseline architecture used for analysis is the ResNet-34.¹³ ResNet-34 is pretrained on ImageNet, a large-scale image database that consists of more than 14 million labeled images. Pretraining optimizes the CNN's layers to perform fundamental object recognition tasks such as edge detection by means of encoding this capacity within the weights of the CNN.¹⁴

The CNN model was trained using the transfer learning approach. Using the established weights and biases of a pre-trained CNN model, it is repurposed to perform a new classification task. The pre-trained CNN model was then trained to predict road quality based on IRI rating data and the corresponding satellite imagery of each road segments through the process called fine-tuning. The fine-tuning process is restricted to the last layer of the CNN, responsible for the classification.

Image classification is performed for two tasks: (i) a four-class classification task corresponding to the IRI ratings (bad, poor, fair, good); and (ii) a two-class classification task corresponding to bad and good, where bad and poor, and fair and good are merged. In total, the 2% cloud cover and 90-day image window restrictions result in a total of 94,000 valid downloaded image

¹¹ Analysis of the three models is performed within the fast.AI framework, which contains modules that allow for rapid programming of neural networks.

¹² The choice of loss function depends on the type of prediction problem. Mean squared error and mean absolute error loss functions are commonly used for regressions, while cross-entropy and hinge loss functions are the most appropriate for binary classification problems.

¹³ ResNets were developed to address the degradation problem of deep CNN, which suggests that difficulties can be experienced in approximating identity mappings by multiple nonlinear layers. Further, ResNet eases the optimization process by providing faster convergence at the early stage. Resnet-34 is a 34-layer CNN version of the ResNet. This is a model that has been pretrained on the ImageNet dataset (He et al. 2015).

¹⁴ The amount of data necessary to train a model is generally an increasing function of the number of model parameters. For a model with P parameters, a reasonable lower bound is 10P datapoints (Nasir and Sassani 2021). For deep learning networks, training datasets for specific problems are not able to meet this requirement. However, the model is able to take advantage of transfer learning.

segments that are available as inputs for the CNN. A cross-validation approach was used to tune hyperparameters and measure the performance of the model. Table 4 gives results of running the analysis for different sample sizes. The analysis is run for 10 epochs of each model. Results show a classification accuracy of about 60% for the binary classification of roads, relative to a classification accuracy of up to 39% for the classification of roads into four classes. The CNN analysis is not completed on the full set of data because of inadequate computational power.¹⁵ However, the analysis by sample shows there is no significant improvement in accuracy based on the size of the dataset.

Table 4: Convolutional Neural Network Results by Sample Size
(Decimal accuracy)

Sample Size	Binary	Four-Class
1,000	0.61	0.37
5,000	0.57	0.37
10,000	0.61	0.38

Source: Authors' calculations using international roughness index data from the Philippine Department of Public Works and Highways and daytime imagery from the Sentinel-2 satellite (accessed 15 August 2022).

Robustness tests are given for various CNN architectures. The architectures included are ResNet-50¹⁶, AlexNet, VGG, and SqueezeNet. These are architectures used for object detection, which are comprised of varying number of layers and use different strategies to deliver their purpose. This test was conducted to determine if the architecture used would have a significant effect on the accuracy of the model. However, as shown in Table 5, there were no significant improvements observed using any of the architectures.

Table 5: Convolutional Neural Network Results by Architecture
(Decimal accuracy)

CNN Architecture	Binary	Four-Class
ResNet-34	0.57	0.37
ResNet-50	0.59	0.36
AlexNet	0.57	0.37
VGG-11	0.60	0.36
SqueezeNet	0.59	0.39

¹⁵ The analysis is run on Google Colab, which offers access to a graphics processing unit. The creation of the road segments dataset is challenging as it requires batch processing from the Google Earth Engine. Processing of thousands of batch tasks, which require clipping and saving operations, is a lengthy operation. Optimization of the process requires clipping of larger road sections, and local processing and clipping of road segments based on the geo-coordinates of each road section's bounding box. An additional challenge is the processing of a large number of road segments within a CNN. Within a Google Colab environment, the computational capacity is insufficient to process close to 100,000 images, even though each image is a few kilobytes in size.

¹⁶ ResNet-50 is a CNN that is 50 layers deep. Instead of using two-layer blocks like ResNet-34, it uses three-layer bottleneck blocks resulting in a higher accuracy.

CNN = convolutional neural network, ResNet = residual neural network.

Source: Authors' calculations using international roughness index data from the Philippine Department of Public Works and Highways and daytime imagery from the Sentinel-2 satellite (accessed 15 August 2022).

2. Tabular Model Results

The tabular model is implemented for 104,000 observations, which correspond to the observations where it was possible to obtain corresponding remotely sensed and modeled data.¹⁷ About 20,000 observations were dropped from the initial sample of 124,276 road segments because of unavailable remotely sensed data. The baseline implemented model included a tabular neural network with three hidden layers, each corresponding to 400, 200, and 100 neurons.

For robustness, the analysis was also run on a combined dataset composed of information on road surface type (classified according to asphalt, concrete, earth, gravel, or surface treatment) and pavement type (including a variety of asphalt and concrete categories, as well as roads with no pavement), and merged with the aggregated remotely sensed data for each road segment. This combined dataset resulted in a total of 82,928 observations. Further, robustness tests were run on a deeper neural network architecture that included five hidden layers, each corresponding to 1,600; 800; 400; 200; and 100 neurons.

The performance of the baseline model shows an accuracy of about 60% for the four-class classification problem, and an accuracy of about 75% for the binary classification problem. The results do not show significant differentiation based on the number of hidden layers, with the performance similar for both the three- and five-layer networks. Note that unlike CNNs, which are pretrained on ImageNet data and fine-tuned to the road segment data, this model is fitted on the tabular dataset. A sample of 80% is used for training, and the remaining 20% is used for testing.

However, the remotely sensed dataset shows better performance than the combined dataset, showing that the inclusion of additional information on surface type and pavement type does not have a significant impact on model performance. In all cases, the remotely sensed data model has a better performance by about two percentage points. The number of observations used in the analysis therefore has a greater impact on model performance.

Table 6: Tabular Neural Network Results by Architecture
(Decimal accuracy)

Tabular Data	Binary		Four-Class	
	Three Layers	Five Layers	Three Layers	Five Layers
Remotely sensed	0.75	0.74	0.60	0.60
Combined dataset	0.73	0.73	0.58	0.58

¹⁷ From this point forward “remotely sensed and modeled data” will be collectively termed as “remotely sensed data.”

Notes:

1. “Remotely sensed” data pertains to aggregated remotely sensed and modeled data for each 100-meter road segment.
2. “Combined dataset” is composed of information on road surface type, pavement type, and remotely sensed and modeled data.

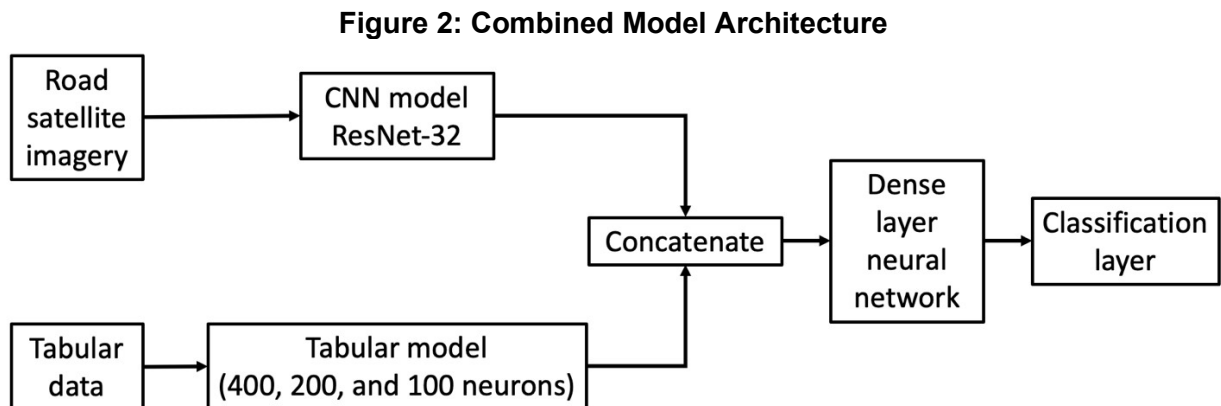
Source: Authors’ calculations using (i) international roughness index data from the Philippine Department of Public Works and Highways; (ii) average daily temperature, and average total precipitation from the European Centre for Medium-Range Weather Forecasts (accessed 09 September 2022); land gradients using the Shuttle Radar Topography Mission of the National Aeronautics and Space Administration (NASA) (accessed 09 September 2022); and total population from WorldPop (accessed 09 September 2022).

3. Combined Model Results

The combined model incorporates both visual and tabular data through an architecture that allows the utilization of both types of data. The architecture used borrows from the integrated model used in the skin melanoma classification challenge organized by the Society for Imaging Informatics in Medicine and International Skin Imaging Collaboration (Ha, Liu, and Liu 2020). This classification problem entailed merging images of skin lesions and patients’ data such as gender, age, and the location of the malignant melanoma, which were then fed into a CNN model and a fully connected neural network to generate final predictions.

The model architecture utilized includes the following (Figure 2):

- (i) a ResNet-34 CNN pretrained on the ImageNet database;
- (ii) a tabular neural network with three hidden layers, each consisting of 400, 200, and 100 neurons; and
- (iii) a combined neural network that takes the concatenated output of the two models and is trained to classify road quality.



CNN = convolutional neural network, ResNet = residual neural network.

Source: Authors’ visualization.

Results from this network have an accuracy of 45% for the four-class model, whereas the binary classification model has an accuracy of 75%. These results are obtained from a random sample of 5,000 observations and are a slight improvement from the CNN model and the tabular model

at the same sample size. Computational constraints do not allow processing of the full set of 94,000 observations. However, the utilization of sample sizes from 1,000 to 10,000 observations does not show a significant improvement in the results.

IV. Discussion

The use of satellite data offers significant opportunities to contribute to improvements in road maintenance. Conventional methods of detecting road quality are expensive and therefore only infrequently implemented. This means that deteriorating roads are frequently not detected early enough, leading to significantly higher eventual costs of road maintenance. It is estimated that the eventual maintenance costs for neglected roads rise to 6 times after 3 years, and up to 18 times after 5 years of neglect (Burningham and Stankevich 2005). Beyond the issue of timely road quality detection, road maintenance is typically underfunded, with estimates that countries spend only 20%–50% of what they should be spending on road maintenance. Therefore, any methods that lower the cost of road quality detection or road repair have great potential to improve overall road maintenance, and thereby allow the realization of the full network effects that are enabled by efficient road networks.

Satellite image data and artificial intelligence offer promise in monitoring the quality of roads. This paper utilizes remotely sensed data derived Google Earth Engine, which offers a combination of visual data on roads at a resolution of 10 meters per pixel (in addition to temperature, precipitation, terrain, and population density data) to detect the quality of roads using CNNs, neural networks using tabular data, and combined networks that utilize both visual data on road segments and tabular data. The results show that tabular data offers potential for the preliminary identification of road segments that are likely to require maintenance, with classification accuracy of up to 75%. While the classification accuracy using these data sources is above 87%, the accuracy is not sufficient to allow full automation of road quality detection.

The analysis offers multiple promising avenues for additional research. Significant resources were expended in optimizing algorithms to segment road sections efficiently, and computational constraints did not allow utilization of the full set of image data. However, information from tabular resources shows that there are improvements from larger tabular datasets. Therefore, leveraging more survey data from multiple countries may offer the possibility of improving performance, particularly given that transfer learning is not used in a tabular neural network.

Further research is also warranted on the following three key topics. First, it is instructive to assess how incorporation of higher-resolution imagery and/or use of super-resolution techniques can potentially enhance the predictive power of the model. Super-resolution refers to the use of machine learning to clarify, sharpen, and upscale the image without losing its content and defining characteristics. Second, it might also be interesting to assess whether an algorithm trained using a dataset from one country may also be applied to produce accurate sample predictions for other countries. This will determine if the algorithm will be adoptable and still have acceptable accuracy rates given the difference in conditions and environments in other countries. Further, the results of such a study may be used to reflect on whether the concept of roughness of roads is universal.¹⁸ Third, it could be useful to explore the possibility of predicting the remaining service life of roads

¹⁸ The authors did a quick assessment by producing out-of-sample predictions for select roads in Thailand using the algorithm trained with data from the Philippines and comparing the predictions with actual roughness of road data. Although the results are encouraging (as the models accurately predicted the qualitative classification of the roads considered), further examination using a larger sample is needed.

following the studies conducted by Ziari et al. (2016) and Al-Suleiman and Shiyab (2003). This study should consider several factors such as pavement type, date of construction or last rehabilitation, traffic volume, and other environmental and external factors, among others.

Nevertheless, the results presented here underscore the potential of satellite imagery-based methods as an alternative way of collecting data on road conditions. Furthermore, by leveraging satellite images, national statistical systems that are responsible in compiling Sustainable Development Goal indicators may also find this approach useful in the context of making the Rural Access Index (Sustainable Development Goal Indicator 9.1.1) widely available at more granular levels.

References

- Abdelaziz, N. et al. 2020. International Roughness Index Prediction Model for Flexible Pavements. *International Journal of Pavement Engineering*. 21 (1). pp. 88–99. <https://doi.org/10.1080/10298436.2018.1441414>.
- Al-Suleiman (Obaidat), T.I. and A.M.S. Shiyab. 2003. Prediction of Pavement Remaining Service Life Using Roughness Data—Case Study in Dubai. *International Journal of Pavement Engineering*. 3 (2). pp. 121–129. <https://doi.org/10.1080/10298430310001634834>.
- Asian Development Bank. 2017. *Meeting Asia's Infrastructure Needs*. Manila. <https://dx.doi.org/10.22617/FLS168388-2>.
- Bashar, M. and C. Torres-Machi. 2021. Performance of Machine Learning Algorithms in Predicting the Pavement International Roughness Index. *Transportation Research Record*. 2675 (5). pp. 226–237. <https://doi.org/10.1177/0361198120986171>.
- Bennett, C., H. De Solminihac, and A. Chamorro. 2006. *Data Collection Technologies for Road Management*. Transport Notes Series; No. TRN 30., World Bank, Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/11776>.
- Burningham, S. and N. Stankevich. 2005. Why Road Maintenance is Important and How to Get It Done. *Transport Notes Series*. 4. Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/11779>.
- Cadamuro, G. A., Muhebwa, and J. Taneja. 2018. Assigning a Grade: Accurate Measurement of Road Quality Using Satellite Imagery. *arXiv*. <https://doi.org/10.48550/arXiv.1812.01699>.
- Copernicus Climate Change Service. 2017. *ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate*. Accessed 09 September 2022 <https://cds.climate.copernicus.eu/cdsapp#!/home>.
- Copernicus Sentinel Data. 2021. Processed by the European Space Agency.
- Department of Public Works and Highways. 2011. Department Order No. 22, Series of 2011. https://dpwh.gov.ph/dpwh/sites/default/files/issuances/DO_022_S2011.pdf
- Department of Public Works and Highways. 2021. DPWH Atlas 2021 (accessed 14 October 2022). https://www.dpwh.gov.ph/dpwh/DPWH_ATLAS/index.htm.
- Eisenbach, M. et al. 2019. Enhancing the Quality of Visual Road Condition Assessment by Deep Learning. *26th World Road Congress*. Abu Dhabi, UAE.
- Farr, T. G. et al. 2007. The Shuttle Radar Topography Mission. *Review of Geophysics*. 45 (2). <https://doi.org/10.1029/2005RG000183>.
- Fiorentini, N. et al. 2021. Can Machine Learning and PS-InSAR Reliably Stand In for Road Profilometric Surveys? *Sensors*. 21 (10). p. 3377. <https://doi.org/10.3390/s21103377>.

- Gaughan, A. et al. 2013. High Resolution Population Distribution Maps for Southeast Asia in 2010 and 2015. *PLOS ONE*. 8 (2). e55882. <https://doi.org/10.1371/journal.pone.0055882>.
- Gerke, V. and C. Heipke. 2008. Image-Based Quality Assessment of Road Databases. *International Journal of Geographical Information Science*. 22 (8). pp. 871–894. <https://doi.org/10.1080/13658810701703258>.
- Gillespie, T. W., Paterson, and M. Sayers. 1986. Guidelines for Conducting and Calibrating Road Roughness Measurements. *World Bank Technical Paper*. WTP 46. Washington, D.C: World Bank. <http://documents.worldbank.org/curated/en/851131468160775725/Guidelines-for-conducting-and-calibrating-road-roughness-measurements>.
- Ha, Q. B., Liu, and F. Liu. 2020. Identifying Melanoma Images using EfficientNet Ensemble: Winning Solution to the SIIM-ISIC Melanoma Classification Challenge. *arXiv*. <https://doi.org/10.48550/arXiv.2010.05351>.
- He, K. et al. 2015. Deep Residual Learning for Image Recognition. *arXiv*. <https://doi.org/10.48550/arXiv.1512.03385>.
- Iandola, F. et al. 2006. SqueezeNet: AlexNet-level Accuracy with 50x Fewer Parameters and <0.5MB Model Size. *arXiv*. <https://doi.org/10.48550/arXiv.1409.1556>.
- Krizhevsky, A. I., Sutskever, and G. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. *25th International Conference on Neural Information Processing Systems - Volume 1*. New York: Curran Associates Inc. pp. 1097–1105.
- Leduc, E. and G. Assaf. 2022. Road Visualization for Smart City: Solution Review with Road Quality Qualification. *Internet of Things*. 12 (100305). <https://doi.org/10.1016/j.iot.2020.100305>.
- Marcelino, P. M., de Lurdes Antunes, and E. Fortunato. 2018. Comprehensive Performance Indicators for Road Pavement Condition Assessment. *Structure and Infrastructure Engineering*. 14 (11). pp. 1433–1445. <https://doi.org/10.1080/15732479.2018.1446179>.
- Mnih, V. and G. Hinton. 2010. Learning to Detect Roads in High-Resolution Aerial Images. In K. Daniilidis, P. Maragos, and N. Paragios, eds. *Computer Vision – ECCV 2010. ECCV 2010. Lecture Notes in Computer Science*. Berlin: Springer. pp. 210–223. https://doi.org/10.1007/978-3-642-15567-3_16.
- Nasir, V. and F. Sassani. 2021. A Review on Deep Learning in Machining and Tool Monitoring: Methods, Opportunities, and Challenges. *The International Journal of Advanced Manufacturing Technology*. 115. pp. 2683–2709. <https://doi.org/10.1007/s00170-021-07325-7>.
- Oshri, B. et al. 2018. Infrastructure Quality Assessment in Africa Using Satellite Imagery and Deep Learning. *24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York: Association for Computing Machinery. pp. 616–625.

- Sabottke, C. and B. Spieler. 2020. The Effect of Image Resolution on Deep Learning in Radiography. *Radiology: Artificial Intelligence*. 2 (1). <https://doi.org/10.1148/ryai.2019190015>.
- Simonyan, K. and A. Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv*. <https://doi.org/10.48550/arXiv.1409.1556>.
- United Nations. 2015. *Transforming our world: the 2030 Agenda for Sustainable Development*. A/RES/70/1. New York. https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf.
- Workman R. and K. McPherson. 2021. Measuring Rural Access for SDG 9.1.1. *Transactions in GIS*. 25. pp. 721– 734. <https://doi.org/10.1111/tgis.12721>.
- Ziari H. et al. 2016. Prediction of IRI in Short and Long Terms for Flexible Pavements: ANN and GMDH Methods. *International Journal of Pavement Engineering*. 17 (9). pp. 776–778. <https://doi.org/10.1080/10298436.2015.1019498>.

Application of Machine Learning Algorithms on Satellite Imagery for Road Quality Monitoring

An Alternative Approach to Road Quality Surveys

This study examines the feasibility of using satellite imagery and artificial intelligence to develop an efficient and cost-effective way to determine and predict the condition of roads. A preliminary algorithm was created and validated using medium-resolution satellite imagery and existing road roughness data from the Philippines. The study used three neural network architectures: the convolutional neural network, tabular, and combined models.

About the Asian Development Bank

ADB is committed to achieving a prosperous, inclusive, resilient, and sustainable Asia and the Pacific, while sustaining its efforts to eradicate extreme poverty. Established in 1966, it is owned by 68 members—49 from the region. Its main instruments for helping its developing member countries are policy dialogue, loans, equity investments, guarantees, grants, and technical assistance.

