Probabilistic Management of Pavement Defects with Image Processing Techniques

Felix OBUNGUTA[1], Kakuya MATSUSHIMA[2] and Junichi SUSAKI[3]

[1]Ph.D. Candidate, Kyoto University, Graduate School of Engineering, Department of Urban Management, Katsura Campus Nishikyo-ku, Kyoto, Japan, 615-8540; E-mail[1]: obunguta.felix.68w@st.kyoto-u.ac.jp/obongutafelix@gmail.com
[2]Associate Professor, Kyoto University, Graduate School of Engineering, Department of Urban Management, Katsura Campus Nishikyo-ku, Kyoto, Japan, 615-8540; E-mail[2]: matsushima.kakuya.7u@kyoto-u.ac.jp
[3]Professor, Kyoto University, Graduate School of Engineering, Department of Urban Management, Katsura Campus Nishikyo-ku, Kyoto, Japan, 615-8540; E-mail[2]: susaki.junichi.3r@kyoto-u.ac.jp

Abstract: Pavement management decisions have traditionally been made by engineers (human-based). However, the pavement stock has recently increased yet management expert numbers are reducing, posing a challenge of how to manage road infrastructure with fewer resources efficiently. Human-based methods are prone to errors that compromise analysis and decisions. More efficient computer-based techniques could offer viable solutions. This study builds a pavement management model with a safety metric output and applies sensitivity analysis to evaluate the most influential inputs from image processing. The study explored image processing techniques considering a trade-off between processing cost and output accuracy. The robustness of the model was tested by comparing its output with the judgement of expert engineers on pavement safety level and its applicability was empirically shown for select roads in Japan.

Keywords: *Pavement management, image processing, sensitivity analysis*

1. INTRODUCTION

(1) Pavement management

Pavement management decisions may be based on the predicted performance of the pavement structure. Infrastructure performance models can be placed into three broad categories; i.e., stochastic (probabilistic), deterministic and computer techniques (Tsuda et al. 2006, Kobayashi et al. 2010, Tabatabaee and Ziyadi 2013, Pérez-Acebo et al. 2019, Obunguta and Matsushima 2020). The Bayesian approach has also been used to improve the prediction of infrastructure performance through updating whenever more data is available (Kobayashi et al. 2012, Tabatabaee and Ziyadi 2013). Stochastic and deterministic techniques may require a minimum of two-point data to predict the performance of infrastructure systems; however, cases of incomplete data including one-point may occur due to a lack of resources such as human and equipment to carry out surveys. Additional data may be generated through multiple imputation (Rubin 1976, 1987) and/or computer techniques could be used to process one-point data and output useful information to support management decisions (Maeda et al. 2018, Zou et al. 2022). Lethanh and Adey (2012) applied the improved stochastic hidden Markov model to model pavement deterioration in case of incomplete monitoring data.

Infrastructure asset management is heavily dependent on infrastructure condition which requires significant amounts of data. In the past, data had been collected by engineers through periodic inspection which is prone to a number of errors such as miss-reporting, omission and/or wrong data entries. Human-based detection and measurement of defects is a highly subjective process liable to bias. The use of expensive specialized damage measurement equipment may not be feasible in some settings and may also disrupt normal traffic flow. A shortage of experts has also resulted in less inspection coverage (Maeda et al. 2018). Furthermore, the collected data is normally manually sorted by a data analyst to eliminate unusable data, a process that may introduce additional errors. The poor data problem is further augmented at the data cleaning stage, where a lot of data is eliminated affecting the power of estimates obtained from prediction models; which blurs management decisions (Obunguta and Matsushima 2020). Accurate and effective computer-based infrastructure management using fewer resources (both human and material) could thus be desirable.

Pavement management decisions are made to minimize costs typically Life Cycle Costs (LCC) of the pavement for a projected period of operation. Kobayashi et al. (2013) developed a pavement management model that optimized inspection and repair for pavements by minimizing LCC. Obunguta and matsushima (2020) optimized the LCC of a pavement system by considering different management policies; i.e., time-dependent and condition-dependent, and explored the effect of preventive maintenance on LCC. Pavement intervention may also be determined by optimizing road usage and utility (Lin and Lin 2011, Liu and Wang 2016, Mizutani et al. 2020).

(2) Sensitivity analysis

Sensitivity Analysis (SA) is a method applied to examine how changes in inputs influence outputs mainly due to the uncertainty surrounding the true values of the inputs (Oakley and O'Hagan 2004). Many engineering problems may involve complex models with a high number of input variables. SA may be performed to identify and prioritize the most influential inputs inorder to reduce the complexity and computational cost of the problem (Iooss and Lemaître 2015, Antoniadis et al. 2021). Pianosi et al. (2016) presented a comprehensive review of Probabilistic Sensitivity Analysis (PSA) and its application

to environmental modelling. Many fields including engineering, economics, environmental science and health have applied PSA to examine the influence of inputs for their large process models. PSA may be local or global. Local sensitivity analysis (LSA) indicates how the output may change if the base input values are slightly perturbed, whereas Global Sensitivity Analysis (GSA) examines the influence on the output by more substantial changes in the inputs. LSA techniques may not be suitable for complex systems because they disregard uncertainities in unchanged input variables. GSA techniques may include correlation and regression analysis, regional sensitivity analysis, density-based, and variance-based methods which include ANOVA and Sobol's method (Sobol' 1993, Pianosi et al. 2016). ANOVA uses F-tests to assess significance but may fail where no true random error exists (Zamanian et al. 2021). Antoniadis et al. (2021) discussed GSA using random forests as an efficient non-parametric method with the advantages of ease of operation for regression problems, and its ability to implicitly deal with high dimension data. Oakley and O'Hagan (2004) suggested a Bayesian approach to PSA which performed effectively using far smaller numbers of model runs than Monte Carlo methods. Zamanian et al. (2021) applied GSA to identify the key variables that infuence crack formation in buried concrete sewer pipes considering uncertainities associated with material properties and loading. Mizutani et al. (2020) applied SA to evaluate how risk control levels and the distance between sections affected repair costs. SA may be applied in pavement engineering to evaluate the most influential inputs applied to estimate outputs for effective decision making so as to prescribe appropriate interventions.

(3) Image processing techniques

Recent technological advancements have led to the development of comparatively lower cost but high quality smartphones which has resulted in the production of abundant smartphone road image data. The images may be stored in datasets such as ImageNet (Deng et al. 2009) and PASCAL VOC (Everingham et al. 2015), CamVid (Cambridge University 2021) and Road Damage Dataset – 2020 (RDD-2020) (Arya et al. 2020a).

For infrastructure systems such as road networks, images are normally collected by taking photos through the car windshield using a smartphone mounted on the dashboard (**Fig. 1**). Car windshield images are complex because they contain a lot of noise (many objects) and are in perspective view. The images may additionally be affected by weather; e.g., lighting and shadows. Plan view images, taken directly above the road surface; for example, using a drone may be simpler but are legally prohibited in many jurisdictions.

Image processing has been applied to many fields such as forestry to evaluate the impacts of policies addressing deforestation (The Mathworks Inc. 2021), transportation infrastructure for road damage detection (Maeda et al. 2018, Arya et al. 2020a, 2021), and dermatology to determine the severity of skin cancer (Kinyanjui et al. 2019) and skin lesions (Mirikharaji et al. 2021). A study by Zou et al. (2022) applies deep learning using the You Only Look Once v4 (YOLOv4) algorithm to detect defects in structures after an earthquake disaster. Maeda et al. (2018) developed a road damage detection system based on the YOLO algorithm using smartphone images in Japan. Thuyet et al. (2022) developed an autonomous road inspection system using deep learning and data obtained using a Laser Crack Measurement System to detect cracks and patches. Other studies such as Goncalves and Givigi (2016), Hong et al. (2020) have developed methods to detect and measure crack defects in civil infrastructure from simple image data, containing few objects.

Fig. 1. Setup of smartphone in car (Arya et al. 2021).

Object recognition systems can be broadly divided into three groups. First, human-based methods, where an inspector observes and measures defects using traditional measurement equipment. Second, microscopic inspection using specialized tools; and third, machine vision, in which defects are identified and quantified automatically by image analysis. Within machine vision, Artificial Neural Networks (ANNs) and convolutional neural networks (CNNs); and pattern recognition using colour models have emerged as the most popular (Goncalves and Givigi 2016). Colour models were an advancement of simple threshold segmentation (e.g., Otsu 1979). Other segmentation methods including graph-based segmentation using the lazysnapping technique and region growing from a seed point have been developed (The Mathworks Inc. 2021).

In deep learning, algorithms built using region proposals and CNNs (R-CNN) have shown high accuracy. The Mask R-CNN algorithm (He et al. 2018); an advancement of Faster R-CNN (Ren et al. 2015), Fast R-CNN (Girshick 2015) and R-CNN (Girshick et al. 2014), is the current state-of-the art algorithm in the family of object detection and segmentation algorithms using region proposals. Mask R-CNN extended Faster R-CNN by adding a branch for predicting segmentation masks from each Region of Interest (RoI) and also replaced the RoIPool layer with the RoIAlign layer that is quantization-free which solved the misalignment challenge in earlier algorithms.

Object detection techniques build a bounding box based on the object category as shown in **Fig. 2** and therefore do not provide information about the size and shape of defects. Image detection and segmentation algorithms such as Mask R-CNN provide a pixel-wise mask for an object which gives more details about its shape and size, and may also be more suitable for segmentating complex images containing overlapping objects, different colours, textures, contrasts and light intensities. The severity of defects, obtained through quantification, is important for the asset management decision process especially considering safety.

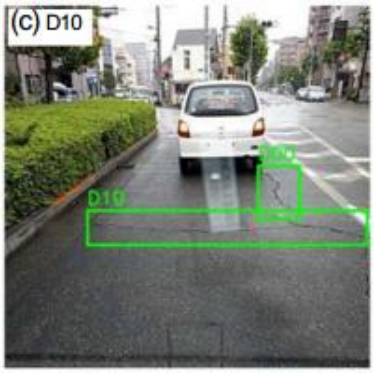


Fig. 2. Crack detection (Maeda et al. 2018).

(4) Image Annotation

Annotation is a vital preliminary step, depite being labour-intensive, before training a deep learning model and therefore should be done as efficiently and as accurately as possible. For object detection, bounding boxes and object labels are manually added to the images at every instance an object is identified by the annotator. For object segmentation, a pixel-wise mask and object label are manually added to the images at each instance. Higher quality and more precise annotations may increase accuracy, however, they require a higher time cost to achieve. More informative annotations may involve detailed manual boundary drawings for a feature of interest, whereas less informative approximate annotations (e.g., bounding boxes or simplified polygons) may require simpler drawings. Therefore, a trade-off may exist between the quality and time cost of annotating images (Mirikharaji et al. 2021). For infrastructure performance evaluations, the accuracy requirements for measurements may not be as strict compared to fields such as health because classification of the defect level need only fall within a specified range. With the accuracy-time cost trade-off, practitioners may decide the needed annotation quality for specific purposes.

Past studies including Greenwald et al. (2022) have attempted to optimize the time cost for annotating images by combining expert, crowd and computer input while ensuring required accuracy levels are met. Xu et al. (2021) applied partial annotation to leverage the advantages of using annotated and unannotated regions in the training process for crowd counting tasks.

(5) Study objectives

The main objective of this study is to explore the possibility of ariving at sound pavement management decisions with minimal human dependence. Specifically, the objectives are:

1) To develop a probabilistic pavement management model based on safety.
2) To examine how changes in model inputs affect the outputs using SA.
3) To empirically show the applicability of the model using deep learning output from the processed RDD-2020.

As far as the authors know, no other study empirically shows the applicability of image processing outputs as inputs for an asset management model including examining the sensitivity of the model inputs and annotation quality. The rest of this article is organised as follows. The Section 2 develops the probabilistic asset management model; Section 3 presents the empirical model application including deep learning on the RDD-2020; Section 4 concludes the article and suggests possible future work.

2. PROBABILISTIC ASSET MANAGEMENT MODEL

(1) Model definition and overview

Consider that a road pavement section $k(k = 1,2, \dots, K)$ has defect density d_k^n estimated from processed image data with n indicating the class of the defect; e.g., alligator cracks, and potholes. Based on the severity of d_k^n with $n(n = 1,2, \dots, N)$, road managers may propose the appropriate intervention A on a section with options; do nothing A_0, sealing or patching A_1, overlay A_2, and reconstruction A_3. The model input variables d_k^n may contain uncertainty that could be due to image quality uncertainties such as lighting conditions, scale for instance perspective view, shadows, and/or pavement infrastructure uncertainties including materials, loading and environmental conditions, and/or processing methods. In this study, processing methods including annotation are emphasized. Each section k, has an input vector $\boldsymbol{d_k} = [d_k^1, \dots, d_k^N]$ that generates an output, the safety metric $S_k = f(\boldsymbol{d_k})$. The priority for intervention on a section k is determined based on the magnitude of the safety metric on that particular section in comparison to other sections and the intervention is proposed following an intervention matrix. The defect densities used to estimate S_k are obtained probabilistically considering different annotation cases i and Intersection over Union (IoU) threshold. This is illustrated in **Fig. 3** with $S_{k=1} > S_{k=2}$, which implies that worse section $k = 1$ should receive priority for intervention.

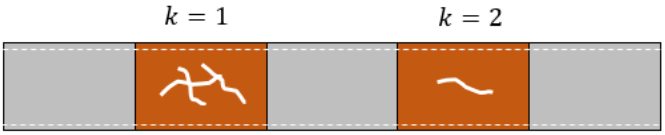


Fig. 3. Illustration of two pavement sections.

(2) Safety metric

Consider that the quality of annotations for images applied in deep learning varies. Assume that the defect density d_k^n for a given annotation case i and IoU can be estimated. The probability $\in [0,1]$ of the defect density falling within an intervention range $[d_{min}^n, d_{max}^n]$ can be expressed as;

$$\Pr\{d_{min}^n < d_k^n \leq d_{max}^n | i, IoU\} = p_k^n(i, IoU) \qquad (1)$$

For a given annotation case and IoU to be acceptable to effectively detect and quantify a defect, the probability $p_k^n(i, IoU)$ should not be less than a set limit p_0^n.

$$p_k^n(i, IoU) \geq p_0^n \qquad \forall n \qquad (2)$$

The annotation case i^* and IoU^* that maximizes the probability of detection of defects across all ranges for a given defect class is obtained by maximizing the objective function;

$$max \prod_{r^n} p_k^n(i, IoU)^{\delta^n} \qquad \forall n, \qquad \forall i, \qquad \forall IoU \qquad (3)$$

$$\delta^n = \begin{cases} 1 & if\ \ IoU > 0 \\ 0 & otherwise \end{cases}$$

Where r^n denotes all ranges for a given defect class n and δ^n is a dummy variable.

The defect densities d_k^{n*} obtained for each i^* and IoU^* are used in the calculation of S_k. The safety metric for a given pavement section k can then be defined as;

$$S_k = \beta_1\, d_k^{1*} + \beta_2\, d_k^{2*} + \cdots + \beta_n\, d_k^{n*} \qquad (4)$$

$$\sum_N \beta_n = 1$$

Where (β_1, β_2, …, β_n) are defects weights set by the road agency based on the influence of a specific defect on safety.

(3) Intervention planning

The choice of a section for intervention A is determined by minimizing the severity level of the section safety metric for the entire pavement stock. Mathematically, this can be expressed as;

$$\min_A \sum_{k=1}^{K} S_k \qquad (5)$$

Subject to

$$A \in \mathbf{\Gamma} \qquad (6)$$

The action A proposed to be performed on a section k has to fall in a feasibility set $\mathbf{\Gamma}$. When action is carried out, it is assumed that the defect density $d_k^{n,A*} = 0$. The intervention on a given section is determined per defect class following an intervention matrix similar to one shown in **Table 1**. This type of intervention decision is used by a number of agencies including the Ministry of Land, Infrastructure, Transport and Tourism of Japan (Kubo 2017). The cut off level for each defect class intervention can be varied by a road agency based on their standards.

(4) Pareto frontier

A Pareto frontier may occur where for two or more elements, the S_k value is the same. In this case, other factors may be considered such as the importance factor of a pavement section relative to others. If other factors are insignificant, then the prioritization of intervention for the sections at the Pareto frontier may be done randomly.

Table 1 Intervention matrix (Adapted from Kubo 2017).

Damage type		Intervention			
Cracks (Cr)	Potholes (Pt)	A_0	A_1	A_2	A_3
0 ≤ Cr < 0.2	Pt = 0	o	x	x	x
0.2 ≤ Cr < 0.4	0 < Pt < 0.1	x	o	x	x
0.4 ≤ Cr ≤ 0.6	0.1 ≤ Pt ≤ 0.3	x	x	o	x
Cr > 0.6	Pt > 0.3	x	x	x	o

Note: o means intervention and x means no intervention.

3. EMPIRICAL APPLICATION

(1) Outline of application

In the empirical application, simple object segmentation methods were explored and a deep learning model was trained using the Mask R-CNN algorithm in Python 3.9.1 to detect and quantify defects and road features (RoIs) in the RDD-2020 so as to estimate the defect densities applied in the probabilistic asset management model. To show the practicality of the model, intervention planning was carried out for select roads in Japan.

(2) Image dataset

The RDD-2020 contains images of 600 × 600 pixels for road surfaces approximately 10 m ahead taken using a camera mounted on a vehicle traveling at an average speed of about 40 km/h (about 10 m/s) capturing an image every second. The dataset is heterogeneous and includes images from India, Japan, and the Czech Republic (**Fig. 4)**.

a) b) c)

Fig. 4. Road images from Japan (a), India (b) and Czech (c).

(3) Simple segmentation methods

a) Overview of methods

There are several segmentation techniques that can be applied to extract features of interest from images. For many images, segmentation needs to be done programmatically as opposed to manually because of computational reasons. This study explored graph-based segmentation using the lazysnapping technique and region growing from a seed point. The segmentation algorithms can be developed in MATLAB and looped through images stored in a specified file directory.

For the lazysnapping technique, the initial background and foreground RoIs are user-dependent. After the RoIs are set, the algorithm programmatically classifies other unallocated image pixels as either background or foreground based on a similarity metric. In **Fig. 5**, consider a 5x5 pixel image with a low pixel (dark) foreground and a high pixel (light) background. The foreground (object) region can be segmented out by specifying the RoI with dimensions [xmin,ymin,width,height] and a background RoI with its own dimensions. The RoIs for each group (fore or background) can be as many as neccessary. The lazysnapping formular can then be used to group pixels based on similarity.

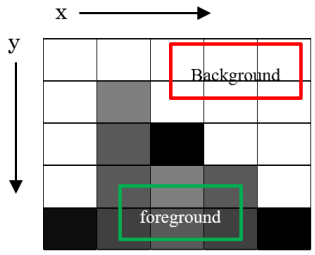


Fig. 5. Graph-based segmentation by lazysnapping with a foreground RoI and a background RoI.

For region growing from seed point(s), the RoI is iteratively grown by comparing all unallocated neighbouring pixels to the RoI based on a similarity measure (Kroon, 2021). This is illustrated in **Fig. 6** were the initial user-dependent seed point Sp with coordinates [x,y] is grown to cover the low pixel object region.

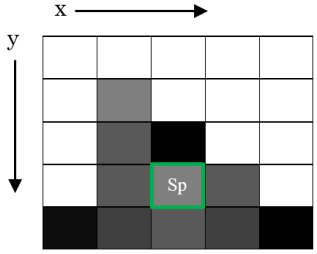


Fig. 6. Region growing from a seed point (Sp).

To run either algorithms programmatically, the RoIs or the Sp need to be pre-set by the user for the entire dataset. This may be inaccurate incase the initial RoIs and Sps do not fall in the pixel area of the feature of interest for all images.

b) Simple segmentation experiments

The segmentation experiments shown in **Fig. 7** and **Fig. 8** highlighted that both lazysnapping and region growing from a seed point were challenged by;

1) Region continuity breakage due to lighting conditions; i.e., dark shadows, bright shiny surfaces and bright reflections probably from the windscreen due to the camera flash.
2) Breakage in segmentation regions due to the colour difference between the road markings such as zebra crossings and lane separations; and the pavement surface.

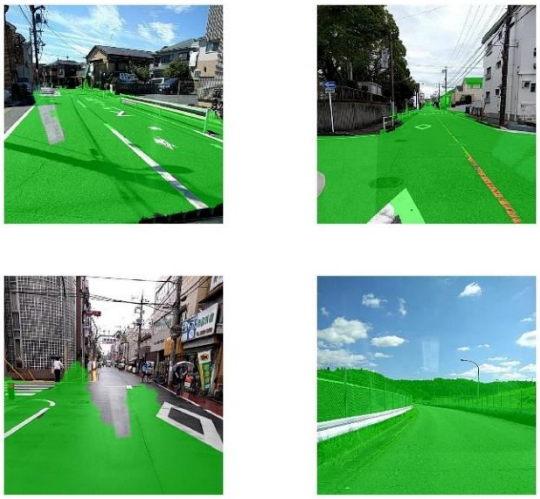


Fig. 7. Segmentation trials using the lazy snapping technique on RDD-2020.

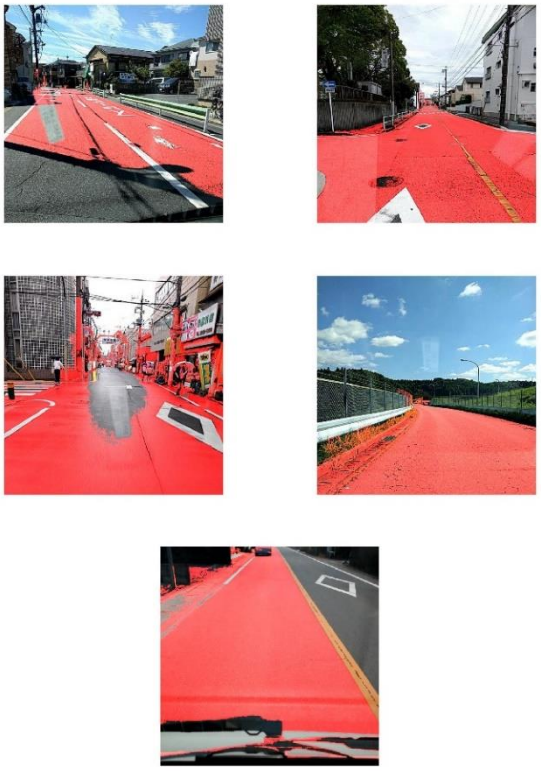


Fig. 8. Segmentation trials using region growing from a seed point on RDD-2020.

The simple segmentation experiments showed that both techniques were challenged by the complexities of segmenting features as a result of lighting and colour changes. This was because segmentation of complex images is challenged by inaccurate initial RoI specification if done programmatically. On the other hand, manual segmentation may be cumbersome, hence, deep learning may offer more accurate segmentation results despite having a higher computational cost (annotation and training) compared to simple segmentation methods.

(4) Deep learning

a) Algorithm

Deep learning involves annotating images then training a model to detect the annotated RoIs. In this study, the deep learning model was trained using the Mask R-CNN algorithm to detect and build a pixel-wise mask on road features and defects. The main steps of the algorithm are detailed in **Table 2** below.

Table 2 Deep learning algorithm.

Algorithm: Deep learning and defect quantification
Start
Step 1: Obtain road pavement images.
Step 2: Sort images.
Step 3: Annotate images in the training and validation set.
Step 4: Input annotated images then train and test the deep learning model.
Step 5: Defect quantification.
Step 6: Output quantified defects; i.e., defect densities.
End

b) Deep learning accuracy

The Mean Average Precision (mAP) is a popular metric in computer vision for evaluating the accuracy of object detectors (Padilla et al., 2020). The measures, precision and recall, are required in the estimation of mAP. Precision is the ratio of true positives to all predicted positives, whereas recall is the ratio of true positives to all actual positives. To explicitly express precision and recall, the following parameters are defined as;

- True Positive (TP): If an object or defect instance is present in the ground truth, and the label and the bounding box of the instance are correctly predicted with Intersection over Union (IoU) ≥ threshold.
- False Positive (FP): If the model predicts an object or defect instance at a particular location in the image, but the instance is not present in the ground truth for that particular image. This also applies to the case when the predicted label doesn't match with the actual label.
- False Negative (FN): If an object or defect instance is present in the ground truth, but the model fails to predict either the correct label or the bounding box of the instance.

The precision and recall are then defined mathematically as;

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

The average precision (AP) is obtained as the average of precision values obtained from the precision-recall (PR) curve for a select set of recall values. The mAP score is the mean of APs over all the object classes, *N*.

$$mAP = \frac{1}{N}\sum_{n=1}^{N} AP_n \quad (9)$$

c) Image annotation

This study explored three annotation cases (**Fig. 9**) with decreasing labour requirements and precision, and compares their accuracy in determining the right defect classifications and quantifications against expert judgements. The study employed the Visual Geometry Group (VGG) Image Annotator (VIA) software to annotate the RDD-2020 images and the annotations were exported in the JSON format. The annotation of the objects of interest was done following **Table 3**. The road feature was added to the defect classes defined by Arya et al. (2020a) with D00, D10 and D20 defining cracks and D40 mainly potholes. **Fig. 10** shows a select road damage image before and after annotation. It took about 2 to 4 minutes for case I, 1 to 2 minutes for Case II and less than 1 minute for Case III; to annotate a single image using human labour depending on the amount of defects observed in the image by the annotator. The **Fig. 11** and **Fig. 12** show the training and validation data statistics, respectively, with a total of 1,165 annotated objects. The low occurrence of D40 defects in Japan compared to the other countries is probably due to better and more regular maintenance.

Fig. 9. Different annotation cases in red, green and blue.

Table 3 Objects of interest.

Object ID	Description
Road	Road surface
D00	Linear crack, longitudinal
D10	Linear crack, lateral
D20	Alligator crack
D40	Pothole, rutting, bump, separation

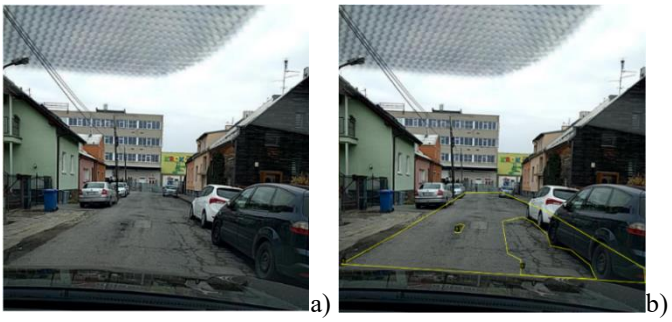
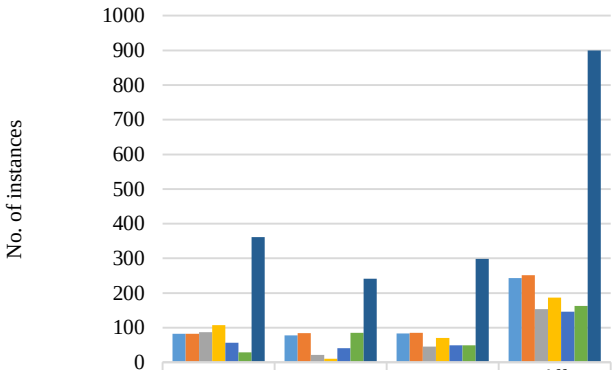
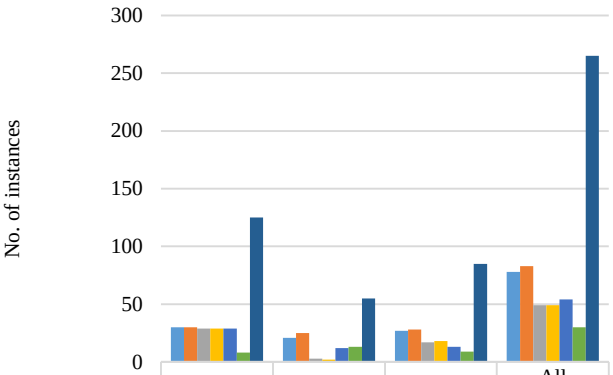


Fig. 10. Road damage image before (a) and after (b) annotation.



	Japan	India	Czech	All countries
No. of images	82	78	83	243
Road	82	84	85	251
D00	87	21	45	153
D10	107	10	70	187
D20	56	41	49	146
D40	29	85	49	163
Total objects	361	241	298	900

Fig. 11. Training data statistics.

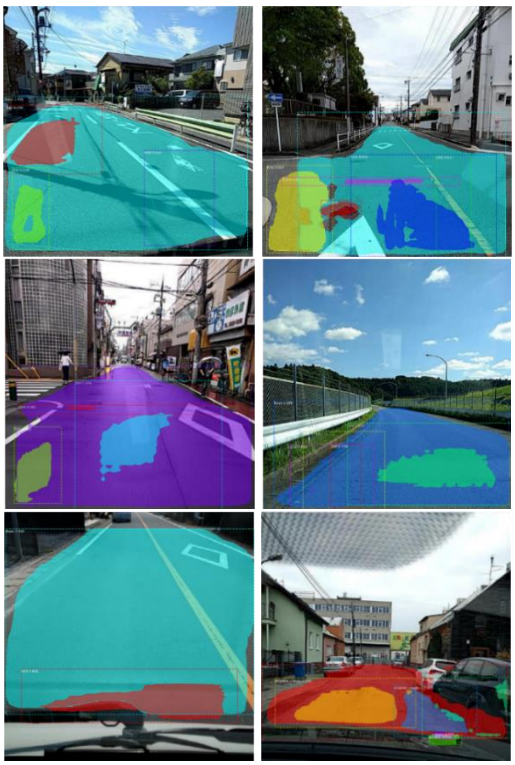


	Japan	India	Czech	All countries
No. of images	30	21	27	78
Road	30	25	28	83
D00	29	3	17	49
D10	29	2	18	49
D20	29	12	13	54
D40	8	13	9	30
Total objects	125	55	85	265

Fig. 12. Validation data statistics.

d) Deep learning experiments

The model was trained for 50 epochs at a learning rate of 0.001 in Python 3.9.1 using an Intel(R) Core(TM) i5-5200U CPU @ 2.20GHz with 4.00GB RAM and 500GB HDD computer. Model accuracy was tested at different IoU thresholds across different object classes considering different annotation precision. The model training took about 33 hours. The **Table 4** shows mAP values per defect class at different IoU thresholds and annotation cases. The **Fig. 13** shows the improved results of the road feature extraction and the defect identification done in parallel without image lighting and colour change inhibitions. The model showed high confidence values of upto 0.99 for prominent road features.



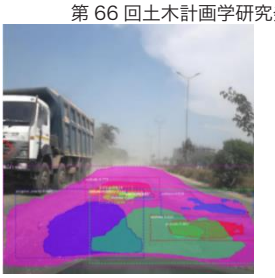


Fig. 13. Detection and segmentation of road features and defects on RDD-2020 for case I.

Comparing across the deferent annotation cases, the mAP increased by an average of 8.3% from Case I to II and by an average of 5.8% from Case II to III considering all IoU thresholds. This is probably because of the increase in the size of the region of interest which results in more detection. However, the difference in detection accuracy is less than 10% which may show the insignificance of annotation precision in generating acceptable defect density estimates for road planning and intervention purposes. Particularly, Case III and II may be competitive because relatively similar mAP levels are achieved at a lower annotation cost compared to Case I. It is also important to note that for a given annotation case and IoU, specific defects may be detected better, for instance, D00 is detected at the highest AP considering annotation case II for all IoU thresholds.

The Road object class had the highest AP values because the road feature is very prominent in all the images which made it easy for the algorithm to learn, detect and segment. On the other hand, the linear cracks consisting of lateral and longitudinal cracks had comparatively lower APs because they had the lowest instances and were generally of smaller size compared to other objects; hence, their detection and segmentation was poorer. As the IoU was decreased, the APs increased across all object classes except for the Road class. This was because the less strict IoU requirement resulted in more object detection. The high AP value for the Road class was stagnant because that was the maximum achievable value.

Table 4 AP per object class at different IoU thresholds.

IoU	Object ID	Case I		Case II		Case III	
		AP	mAP	AP	mAP	AP	mAP
0.7	Road	0.9794	0.3436	1.0	0.3963	1.0	0.4097
	D00	0.0833		0.3280		0.1517	
	D10	0.3556		0.0407		0.4520	
	D20	0.1011		0.2045		0.0824	
	D40	0.1985		0.4081		0.3624	
0.5	Road	0.9794	0.5738	1.0	0.6328	1.0	0.6979
	D00	0.3444		0.6472		0.5992	
	D10	0.5799		0.1512		0.7674	
	D20	0.5337		0.5704		0.4843	
	D40	0.4318		0.7951		0.6386	
0.3	Road	0.9794	0.7020	1.0	0.7284	1.0	0.7645
	D00	0.6083		0.7358		0.7258	
	D10	0.5950		0.2523		0.7681	
	D20	0.7291		0.7893		0.6039	
	D40	0.5984		0.8644		0.7245	
0.1	Road	0.9794	0.7046	1.0	0.7309	1.0	0.7645
	D00	0.6083		0.7358		0.7258	
	D10	0.5950		0.2523		0.7681	
	D20	0.7397		0.8019		0.6039	
	D40	0.6005		0.8644		0.7245	

e) Defects density

The extent of defects could be estimated from the segmented images by calculating the ratio of the size of defect pixels to size of pavement pixels. This was considered because the images may be taken at different perspectives and using different smartphones and hence may have different relative sizes. This estimation of the defects density d_k^n in an image taken at a specific location was done so as to facilitate comparisons between different road sections. The defects density is also similar to the cracking ratio defined by the Japan Road Association (Kubo 2017).

$$d_k^n = \frac{No.of\ defect\ pixels}{No.of\ total\ pavement\ pixels} \tag{10}$$

The defect densities were estimated for 1,660 select sections in Adachi City, Japan. Defect densities greater than 1 were due to the detection and segmentation of multiple object instances, overlap and partial detection of road features in the images. From **Fig. 14**, more severe defects can be detected when the IoU is reduced because a less strict IoU enables more detection as the predicted and ground truth need not overlap much. Also, more serious defects can be detected from annotation case I to II to III for IoU 0.7. This is probably due to better model learning as objects were more precisely defined in case I which increased detection despite the high IoU threshold. However, from IoU 0.5, there is more defect detection for case II compared to other cases. The increase in detection for case III for IoU 0.3 and 0.1 could be attributed to variations in the subjective annotation and probably model learning. This result may be used to empirically show the importance of the IoU threshold and annotation precision in defect detection. The output could encourage the agency to be more proactive in intervention and carry out more detailed tests on the candidate pavement sections.

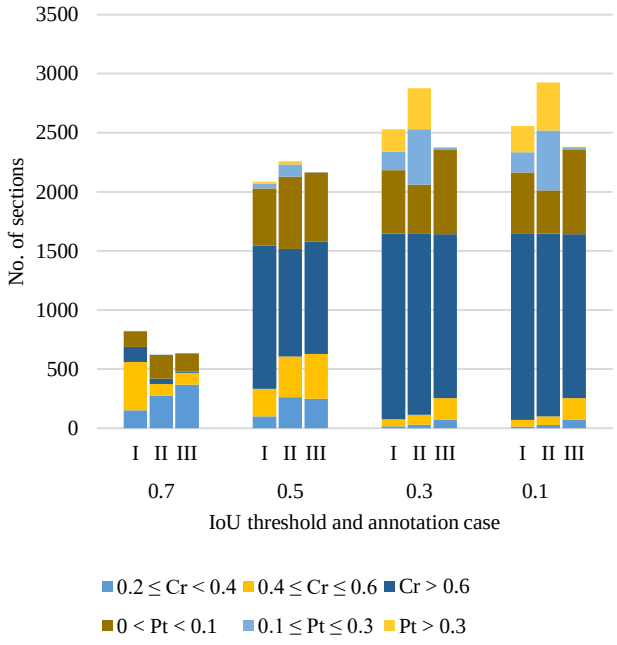


Fig. 14. Defect densities for sections requiring intervention.

(5) Estimation of the safety metric and sensitivity analysis

The choice of annotation case and IoU was obtained by maximizing the objective function for each defect class (choice 1) as shown in **Table 5** with p_0^n set to 0. The less costly annotation case III with IoU of 0.5 and case II with IoU 0.1 were selected to detect cracks and potholes, respectively, based on maximization of the probability and likelihood of detection. This result suggested that the precision of annotations may be insignificant for road defect detection tasks.

Table 5 Probabilistic choice 1 for IoU and annotation case.

Defect range	Objective value	Choice 1 (*i*, IoU)
0.2 ≤ Cr < 0.4	0.0197	III, 0.5
0.4 ≤ Cr ≤ 0.6		
Cr > 0.6		
0 < Pt < 0.1	0.0164	II, 0.1
0.1 ≤ Pt ≤ 0.3		
Pt > 0.3		

The choice of the annotation case and IoU were determined through an optimization process which produced a unique solution that may not necessitate sensitivity analysis. However, the variability of the output safety metric can be evaluated by choosing the second highest (choice 2) and third highest (choice 3) objective value for the same defect ranges. This generates varied annotation cases and IoU threshold choices as shown in **Table 6**. The most costly and most precise annotation case I was not selected until choice 3 which shows the insignificance of annotation precision. The high IoU of 0.7 was not selected for all choices which suggests that very high IoUs may be undesirable for road defect detection tasks.

Table 6 Other Probabilistic choices for IoU and annotation case.

Defects	Choice 2	Choice 3
Cracks	II, 0.5 (0.0180)	I, 0.5 (0.0062)
Potholes	II, 0.3 (0.0148)	I, 0.1 (0.0043)

The values in parentheses are the objective values.

In the empirical application, the defects d_k^1 and d_k^2 were defined to represent cracks and potholes, respectively. The parameters β_1 and β_2 were set to 0.3 and 0.7, respectively. The safety metric can specifically be expressed as;

$$S_k = 0.3\, d_k^{1*} + 0.7\, d_k^{2*} \tag{11}$$

The **Fig. 15** and **Fig. 16** show plots of the safety metric for choice 2 vs. choice 1 and choice 3 vs. choice 1. Based on the gradient and size of the intercept of the lines of best fit, it can be seen that choice 2 varied less from choice 1 compared to choice 3. A bigger perturbation of the input resulted in higher variability of the output safety metric.

The sections with higher S_k values (points furthest from the origin) should receive higher priority for intervention. The proposed intervention for a section can be determined following the intervention matrix after prioritization.

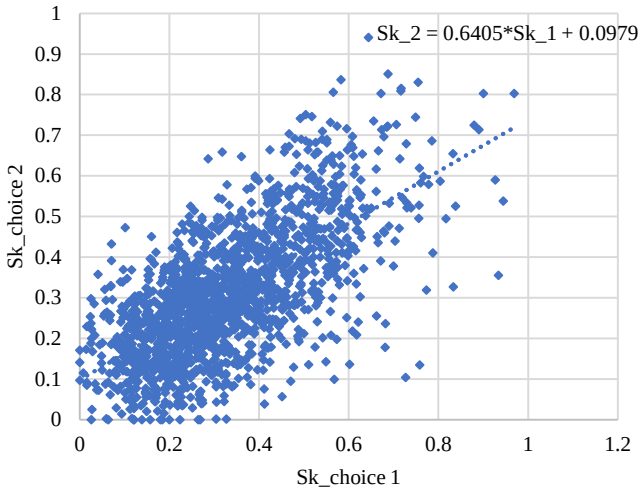


Fig. 15. Safety metric second vs first choice.

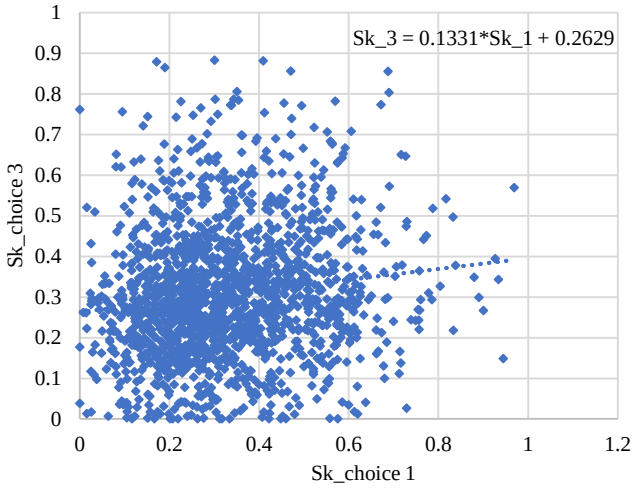


Fig. 16. Safety metric third vs first choice.

(6) Discussion

This study trained a deep learning model using the Mask R-CNN algorithm using the publicly available RDD-2020 database. To minimise the challenge of incorrect object identification, two methods were suggested; i.e., pre-segmentation and model training in succesion, and using deep learning to extract all features of interest in parallel. Pre-segmentation was challenged by segmentation breakage due to lighting conditions and colour changes in the images, and unwanted objects were pre-segmented in several cases. Deep learning in which all objects of interest were identified and segmented in parallel showed more promising results with road features segmented with high AP values. Deep learning made it convenient to estimate the defects densities because defects were obtained in parallel with road features. The study developed an asset management model to prioritise the intervention works for a group of road sections and contained an empirical application to select roads in Japan. The importance of the IoU threshold and annotation case was evaluated and it was shown that low IoU resulted in more detection, where as the annotation precision was insignificant. More detection may result in over estimation of defects, however, this could encourage proactive intervention and further investigations for candidate sections.

4. CONCLUSIONS

This study proposed a framework to feasibly apply deep learning model results to asset management. The study used the publicly available smartphone road image data from Japan, India and the Czech republic to train and validate a deep learning model built on the Mask R-CNN algorithm. The study showed that with fewer resources for management amidst an increasing infrastructure stock, computer vision promises safer and more efficient asset management and planning compared to the current human dependent practice. The study empirically showed the following:

1) The study showed that one-point data obtained from a single image dataset can be efficiently used to support intervention choices on infrastructure.
2) The choice of the IoU threshold significantly affected defect detection and hence should be carefully selected considering safety standards. The annotation precision was insignificant in defect detection and so less costly simplified polygons may suffice.

In future, better asset management models can be developed when more data is consistently obtained because current state-of-the-art asset management models require at least two-point condition data to model deterioration processes and perform LCC analysis. Building algorithms that detect patterns of defects is a possible area for future research because defect patterns could indicate the failure type; for example, diagonal cracks are indicative of shear failure in reinforced concrete structures. The detection and segmentation algorithms also need to be generally improved. A positive feedback loop can also be created between data collection and future asset management needs. For instance, to make the data more usable for planning the GPS coordinates of the photos could be included in the database so that the road sections can be better identified and linked to road network maps. It is recommended that a methodology is developed to make annotations more objective than subjective to minimize variability due to different annotators.

ACKNOWLEDGEMENTS

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI [Grant Number: 21K04288]. The authors appreciate Sekimoto Laboratory at the University of Tokyo for making the road damage dataset publicly available.

DECLARATION OF INTEREST STATEMENT

There was no conflict of interest reported by the authors.

REFERENCES

1) Antoniadis, A., Lambert-Lacroix, S. and Poggi, J.-M.: Random forests for global sensitivity analysis: A selective review, *Reliability Engineering and System Safety*, Vol. 206, 107312, ISSN 0951-8320, https://doi.org/10.1016/j.ress.2020.107312, 2021.

2) Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashiyama, T. and Sekimoto, Y.: Transfer Learning-based Road Damage Detection for Multiple Countries, arXiv preprint arXiv:2008.13101, 2020a.
3) Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashiyama, T. and Sekimoto, Y.: Global Road Damage Detection: State-of-the-art Solutions, arXiv preprint arXiv:2011.08740, 2020b.
4) Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashiyama, T. and Sekimoto, Y.: Deep learning-based road damage detection and classification for multiple countries, *Automation in Construction*, Vol.132, 103935, 2021.
5) Cambridge University: CamVid dataset. http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/. Accessed on December 13, 2021.
6) Deng, J., Dong, W., Socher, R., Li, L. -J., Li, K. and Fei-Fei, L.: ImageNet: A large-scale hierarchical image database, *IEEE Conference on Computer Vision and Pattern Recognition*, Florida, doi: 10.1109/CVPR.2009.5206848, pp. 248-255, 2009.
7) Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J. and Zisserman, A.: The Pascal visual object classes challenge: a retrospective, *International Journal of Computer Vision*, Vol.111, No.1, pp. 98–136, 2015.
8) Girshick, R., Donahue, J., Darrell, T. and Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation, *IEEE Conference on Computer Vision and Pattern Recognition*, Ohio, pp. 580–587, 2014.
9) Girshick, R.: Fast R-CNN, *IEEE International Conference on Computer Vision*, Boston, MA, pp. 1440–1448, 2015.
10) Goncalves, L. R. and Givigi, S. N.: Automatic Crack Detection and Measurement Based on Image Analysis, *IEEE Transactions on Instrumentation and Measurement*, Vol.65, No.3, pp. 583–590, 2016.
11) Greenwald, N.F., Miller, G., Moen, E. et al.: Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning, *Nature Biotechnology*, 40, https://doi.org/10.1038/s41587-021-01094-0, pp.555–565, 2022.
12) He, K., Gkioxari, G., Dollar, P. and Girshick, R.: Mask R-CNN, Facebook AI Research (FAIR), arXiv:1703.06870v3, 2018.
13) Hong, J.W., Jin, S. and Lee, S.E.: A vision-based approach for autonomous crack width measurement with flexible kernel, *Automation in Construction*, Vol.110, 103019, 2020.
14) Iooss, B. and Lemaître. P.: A review on global sensitivity analysis methods. In: Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications, Springer, hal-00975701, 2015.
15) Kinyanjui, N.M., Odonga, T., Cintas, C., Codella, N.C.F., Panda, R., Sattigeri, P. and Varshney, K.R.: Estimating Skin Tone and Effects on Classification Performance in Dermatology Datasets, *NeurIPS Workshop on Fair ML for Health*, Vancouver, Canada, 2019.
16) Kobayashi, K., Do, M. and Han, D.: Estimation of Markovian transition probabilities for pavement deterioration forecasting, *KSCE Journal of Civil Engineering*, Vol.14, No.3, pp. 343–351, 2010.
17) Kobayashi, K., Eguchi, M., Oi, A., Aoki, K. and Kaito, K.: The Optimal Implementation Policy for Inspecting Pavement with Deterioration Uncertainty, *Journal of Japan Society of Civil Engineers*, Vol.1, No.1, pp. 551–568, 2013.
18) Kobayashi, K., Kaito, K. and Lethanh, N.: A Bayesian estimation method to improve deterioration prediction for infrastructure system with Markov chain model, *International Journal of Architecture, Engineering and Construction*, Vol.1, No.1, doi:10.7492/IJAEC.2012.001, pp. 1–13, 2012.
19) Kroon, D.: Region Growing (https://www.mathworks.com/matlabcentral/fileexchange/19084-region-growing), MATLAB Central File Exchange. Retrieved December 13, 2021.
20) Kubo, K.: Pavement Maintenance in Japan, *Road Conference International Symposium*, 2017.
21) Lethanh, N., and Adey, B. T.: A hidden Markov model for modeling pavement deterioration under incomplete monitoring data, *International Journal of Civil and Environmental Engineering*, Vol.6, No.1, pp. 1–8, 2012.
22) Lin, K. and Lin, C.: Applying Utility Theory to Cost Allocation of Pavement Maintenance and Repair, *International Journal of Pavement Research and Technology*, Vol.4, No.4, pp. 212–221, 2011.
23) Liu, H. and Wang, D. Z. W.: Modeling and solving discrete network design problem with stochastic user equilibrium, *Journal of Advanced Transportation*, Vol.50, No.7, pp. 1295–1313, 2016.
24) Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T. and Omata, H.: Road damage detection and classification using deep neural networks with smartphone images, *Journal of Computer-Aided Civil and Infrastructure Engineering*, Vol.33, pp. 1127–1141, 2018.
25) Mirikharaji, Z., Abhishek, K., Izadi, S. and Hamarneh, G.: D-LEMA: Deep Learning Ensembles from Multiple Annotations Application to Skin Lesion Segmentation, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 1837-1846, Held virtually, 2021.
26) Mizutani, D., Nakazato, Y., and Lee, J.: Network-level synchronized pavement repair and work zone policies: Optimal solution and rule-based approximation, Transportation Research Part C: Emerging Technologies, Vol. 120, https://doi.org/10.1016/j.trc.2020.102797, 102797, 2020.
27) Oakley, J. E. and O'Hagan, A.: Probabilistic sensitivity analysis of complex models: A Bayesian approach, *Journal of Royal Statistical Society B*, Vol.66, No.3, pp. 751–769, 2004.
28) Obunguta, F. and Matsushima, K.: Optimal pavement management strategy development with a stochastic model and its practical application to Ugandan national roads, *International Journal of Pavement Engineering*, DOI: 10.1080/10298436.2020. 1857759, 2020.

29) Otsu, N.: A Threshold Selection Method from Gray-Level Histograms, *IEEE transactions on systems, man, and cybernetics*, Vol.9, No.1, pp. 62–66, 1979.
30) Padilla, R., Netto, S. L. and da Silva, E. A. B.: A Survey on Performance Metrics for Object-Detection Algorithms, *Proceedings of the International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 237–242, 2020.
31) Pérez-Acebo, H., Mindrab, N., Railean, A. and Rojí, E.: Rigid pavement performance models by means of Markov Chains with half-year step time*, International Journal of Pavement Engineering,* Vol.20, pp. 830–843. doi:10.1080/10298436. 2017.1353390, 2019.
32) Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., and Wagener, T.: Sensitivity analysis of environmental models: A systematic review with practical workflow, *Environmental Modelling & Software*, Vol.79, doi: 10.1016/j.envsoft.2016.02.008, pp. 214–232, 2016.
33) Ren, S., He, K., Girshick, R. and Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks, *NIPS*, 2015.
34) Rubin D.B., Inference and missing data. *Biometrika*, Vol.63, pp. 581–592, 1976.
35) Rubin D.B., Multiple imputation for non-response in surveys. New York: John Wiley, 1987.
36) Sobol, I.M.: Sensitivity estimates for non-linear mathematical models, *Mathematical Modelling and Computational Experiments*, Vol.1, pp. 407–414, 1993.
37) Tabatabaee, N. and Ziyadi, M.: Bayesian approach to updating Markov-based models for predicting pavement performance, *Transportation Research Record: Journal of the Transportation Research Board*, Vol.2366, No.1, doi:10.3141/2366-04, pp. 34–42, 2013.
38) The MathWorks Inc., 2021. (https://uk.mathworks.com/help/images/ref/lazysnapping.html).
39) Thuyet, D. Q., Jomoto, M., Hirakawa, K., Lei Swe, Y. L..: Development of an Autonomous Road Surface Damage Inspection Program Using Deep Convolutional Neural Network, *Journal of JSCE*, Vol. 10, Issue 1, pp. 235–246, https://doi.org/10.2208/journalofjsce.10.1_235, 2022.
40) Tsuda, T., Kaito, K., Aoki, K. and Kobayashi, K.: Estimating Markovian Transition Probabilities for Bridge Deterioration Forecasting, *Structural Engineering/Earthquake Engineering*, Vol.23, No.2, pp. 241s–256s, 2006.
41) Xu, Y., Zhong, Z., Lian, D., Li, J., Li, Z., Xu, X., and Gao, S.: Crowd Counting with Partial Annotations in an Image, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15570-15579, Held virtually, 2021.
42) Zamanian, S., Hur, J. and Shafieezadeh, A.: Significant variables for leakage and collapse of buried concrete sewer pipes: a global sensitivity analysis via Bayesian additive regression trees and Sobol' indices, *Structure and Infrastructure Engineering*, Vol.17, No.5, DOI: 10.1080/15732479.2020.1762674, pp. 676-688, 2021.
43) Zou, D., Zhang, M., Bai, Z., Liu, T., Zhou, A., Wang, X., Cui, W., and Zhang, S.: Multi-category damage detection and safety assessment of post-earthquake reinforced concrete structures using deep learning, *Journal of Computer-Aided Civil and Infrastructure Engineering*, https://doi.org/10.1111/mice.12815, pp. 1–17, 2022.

(Received September 29, 2022)