IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Mathematical Modelling of Engineering Problems
Vol. 13, No. 2, February, 2026, pp. 334-346
Journal homepage: http://iieta.org/journals/mmep

Spatial Mapping of Pavement Distress Using Digital Image Data and a CNN-Assisted Clustering Framework

Kariyam[1*], Feri Wijayanto[2], Edy Widodo[1]

[1] Statistics Department, Universitas Islam Indonesia, Yogyakarta 55584, Indonesia
[2] Informatics Department, Universitas Islam Indonesia, Yogyakarta 55584, Indonesia

Corresponding Author Email: kariyam@uii.ac.id

https://doi.org/10.18280/mmep.130210

Received: 4 December 2025
Revised: 28 January 2026
Accepted: 6 February 2026
Available online: 15 March 2026

Keywords:
pavement distress, convolutional neural networks, clustering analysis, road condition assessment, spatial mapping, maintenance prioritization

ABSTRACT

Efficient and scalable monitoring of road infrastructure is critical for maintenance prioritization and sustainable transportation planning, especially in resource-constrained environments. This study proposes an Integrative Convolutional Neural Network–Assisted Clustering Framework (ICCF) that connects image-level pavement distress detection with segment-level spatial mapping of road conditions. Rather than introducing new learning algorithms, ICCF emphasizes the operational integration of established Convolutional Neural Network (CNN) models, statistical aggregation, and clustering to support an interpretable, maintenance-oriented analysis. The framework comprises four sequential stages: CNN-based image-level detection and classification of pavement distress, aggregation of predictions into segment-level damage profiles, partition-based clustering, and spatial mapping of road conditions. Five CNN architectures (VGG16, ResNet50V2, ResNet152V2, InceptionV3, and Xception) were evaluated using 9,345 augmented images representing potholes, corrugation, and alligator cracking. ResNet50V2 exhibited the most stable performance, achieving approximately 95% validation accuracy with minimal generalization loss on the test set. Aggregated damage profiles from 108 road segments were then clustered using a block-based k-medoids approach, producing two interpretable road-condition groups associated with moderate and severe damage levels. The resulting spatial cluster maps provide actionable insights that extend beyond simple summary statistics, supporting data-driven prioritization of road maintenance interventions.

1. INTRODUCTION

Monitoring road infrastructure conditions efficiently is crucial for achieving sustainable transportation, strengthening infrastructure resilience, and supporting long-term maintenance planning [1]. Given the essential role of road networks in economic and social activities, timely and accurate pavement condition assessment is necessary to mitigate safety risks and optimize public investment in road rehabilitation [2].

Conventional road inspection practices rely heavily on manual surveys and visual assessments conducted by trained personnel. Although widely adopted, these approaches are often associated with high operational costs, time inefficiency, and subjectivity in damage evaluation, particularly when applied to extensive road networks [3]. Moreover, manual inspections are difficult to scale and may fail to capture rapid changes in pavement conditions caused by traffic loads and environmental factors [4].

The growing availability of digital road imagery acquired from unmanned aerial vehicles, smartphones, and vehicle-mounted cameras has accelerated the adoption of image-based analysis for pavement monitoring [5]. In this context, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have demonstrated strong capabilities in automatically extracting hierarchical spatial features from images. CNN-based models have been successfully applied to detect and classify various forms of road surface damage, such as cracks, potholes, and surface deformations, under diverse lighting and background conditions [6, 7].

Despite these advances, most CNN-based studies primarily focus on improving image-level classification or segmentation accuracy. The resulting outputs are typically limited to localized damage maps and do not provide higher-level aggregation or interpretation at the road-segment or network level [8]. This limitation reduces the practical usefulness of such systems for road authorities, who require summarized and interpretable information to support maintenance prioritization and strategic planning [9].

In parallel, clustering techniques have long been employed in infrastructure management and decision-support systems to group homogeneous entities and reveal latent structural patterns in data [10]. Partition-based methods such as k-means and k-medoids have been applied to classify road segments based on numerical indicators, including traffic intensity, environmental performance, and maintenance-related variables [11, 12]. However, conventional clustering algorithms are not designed to operate directly on high-

dimensional image data and often fail to capture the latent visual characteristics that define similarity among road surface conditions [13].

To address the challenges associated with high-dimensional data, representation learning approaches based on autoencoders have been introduced to project complex inputs into compact latent spaces suitable for unsupervised grouping [14]. While these methods have shown promising results in industrial inspection and remote sensing applications, their adoption in ground-level road imagery and operational road-maintenance workflows remains limited [15].

Motivated by these gaps, this study proposes an application-oriented analytical pipeline termed the Integrative Convolutional Neural Network–Assisted Clustering Framework (ICCF). ICCF integrates CNN-based image-level road damage detection, structured statistical aggregation, and partition-based clustering to support interpretable road-condition mapping. Rather than introducing new learning or clustering algorithms, the framework emphasizes the practical integration of established methods into a unified pipeline designed for real-world deployment. By bridging image-level analysis and network-level interpretation, ICCF aims to provide actionable insights for maintenance planning and to support sustainable infrastructure management.

2. RELATED WORK

Research on road-condition detection and mapping can be broadly categorized into three major streams:

(A) traditional image-processing and statistical approaches,

(B) deep learning–based detection and segmentation, and

(C) representation learning and clustering approaches for infrastructure imagery.

2.1 Traditional image-processing and statistical approaches

Before the widespread adoption of deep learning, pavement inspection relied primarily on classical image-processing techniques such as thresholding, morphological filtering, edge detection, and texture analysis [16]. Although computationally efficient, these methods are highly sensitive to illumination variation, occlusion, and surface irregularities, limiting their robustness in real-world settings. To address these limitations, hybrid approaches combining multiple handcrafted descriptors — such as Local Binary Patterns and Gabor filters — were proposed to improve detection accuracy. However, their performance remained dataset-dependent and difficult to generalize across different road types and image acquisition conditions.

In parallel, clustering analysis has been widely applied to structured numerical and mixed-type infrastructure data to support decision-making processes. Partition-based and hierarchical methods — including k-means, k-medoids, and agglomerative clustering — have been used to classify road segments based on traffic intensity, maintenance cost, or environmental performance indicators [17-20]. Schubert and Rousseeuw [11] demonstrated that k-medoids performance can be substantially improved through refined medoid initialization and optimized swap strategies, reducing computational complexity while maintaining clustering accuracy. Despite their effectiveness for structured indicators, these clustering approaches are not designed to handle high-dimensional image data directly and therefore remain disconnected from image-based road damage detection pipelines.

2.2 Deep learning-based detection and segmentation

With advances in Graphics Processing Unit computing, CNNs have become the dominant paradigm for pavement defect detection and segmentation. CNN-based models have demonstrated strong performance in identifying cracks, potholes, and surface deformations under varying illumination and background conditions [6, 7, 21].

More recent studies have explored advanced architectures, including edge-aware CNNs and multi-scale feature extraction, to enhance crack segmentation accuracy [22, 23]. Transformer-based vision models have also gained attention in pavement analysis. Lu et al. [24] introduced a Swin Transformer–U-Net hybrid that significantly improved segmentation performance in cluttered and noisy environments. Comprehensive reviews highlight a growing trend toward supervised, semi-supervised, and unsupervised deep learning paradigms aimed at reducing annotation costs and improving scalability [25].

Despite these methodological advances, most deep learning–based studies focus primarily on image-level detection or segmentation accuracy. Their outputs are typically limited to pixel-level or image-level predictions and are rarely aggregated or analyzed at the road-segment or network level. Consequently, these approaches offer limited support for strategic maintenance planning and infrastructure-level decision-making.

2.3 Representation learning and clustering for spatial road-condition mapping

Representation learning–based clustering approaches aim to combine feature extraction and unsupervised grouping to uncover latent data structures. Deep embedded clustering and its variants integrate representation learning with clustering objectives to produce compact and discriminative feature embeddings [26]. Extensions incorporating autoencoder architectures have demonstrated improved stability and clustering performance, particularly for high-dimensional data [27].

In industrial inspection and civil infrastructure contexts, unsupervised and weakly supervised representation learning has been shown to reduce dependence on labeled data while improving scalability and operational feasibility [28]. However, applications of representation learning and clustering to ground-level road imagery remain limited. Existing studies predominantly focus on satellite imagery or non-road infrastructure domains, such as land-cover classification and bridge damage detection [29].

Moreover, few studies explicitly link image-based damage detection with segment-level aggregation and network-scale road-condition mapping. This gap highlights the need for integrative frameworks that connect CNN-based image analysis with interpretable clustering and spatial summarization mechanisms. The ICCF framework proposed in this study addresses this gap by emphasizing the operational integration of image-level detection, structured aggregation, and clustering-based analysis for road-condition mapping.

3. THE INTEGRATIVE CONVOLUTIONAL NEURAL NETWORK–ASSISTED CLUSTERING FRAMEWORK SCHEME FOR SPATIAL ROAD DAMAGE MAPPING

3.1 Overview of the Integrative Convolutional Neural Network–Assisted Clustering Framework scheme

This study proposes ICCF as an application-oriented analytical pipeline for spatial road-condition mapping using digital image data. ICCF is not intended as a new learning or clustering algorithm. Instead, it integrates established deep learning–based image analysis with conventional statistical aggregation and clustering techniques into a unified and interpretable workflow.

The primary objective of ICCF is to bridge the gap between image-level road damage detection and segment-level infrastructure mapping, which is essential for maintenance planning at the network scale. While CNN-based models have demonstrated high accuracy in detecting road defects from images, their outputs are often limited to local predictions and are not directly suitable for decision-making by road authorities. ICCF addresses this limitation by transforming image-level predictions into structured road-segment representations and grouping them into meaningful condition categories. The ICCF pipeline consists of four main stages:

(i) CNN-based road damage detection at the image level,

(ii) Feature aggregation at the road-segment level,

(iii) Partition-based clustering assisted by CNN-derived features, and

(iv) Spatial interpretation for maintenance-oriented mapping.

This design emphasizes operational applicability, transparency, and scalability, rather than methodological novelty. The four core stages of the ICCF process are shown in Figure 1.

(i) Road damage image detection using CNN

At the first stage, ICCF employs pretrained CNN architectures to identify road surface damage from digital images. Established CNN models such as ResNet50V2, ResNet152V2, InceptionV3, Xception, and VGGNet-16 are used due to their proven effectiveness in visual feature extraction and classification tasks.

Each model is fine-tuned for multi-class road damage recognition by adapting the output layer to the target damage categories. The CNNs operate at the image level, producing class probabilities that represent the likelihood of different damage types. These outputs serve as descriptive indicators of road condition rather than final decision variables.

Importantly, ICCF does not modify the internal learning mechanisms of the CNNs. The models are used as feature extractors and predictors, providing consistent and repeatable image-level assessments that can be aggregated at higher spatial scales.

(ii) Road-segment feature aggregation

To support spatial road-condition mapping, ICCF shifts the analytical focus from individual images to road segments as the fundamental decision units. Each road segment may contain multiple images captured under varying conditions, such as different viewpoints or illumination levels.

For each segment, CNN outputs are aggregated using descriptive statistical measures, including proportion of predicted damage classification, frequency of severe damage indicators, or variability measures reflecting heterogeneity within the segment. This aggregation step transforms unstructured image-level predictions into structured numerical feature vectors that characterize the overall condition of each road segment. By operating at this level, ICCF aligns the analytical output with the practical needs of infrastructure management, where interventions are planned per segment rather than per image.

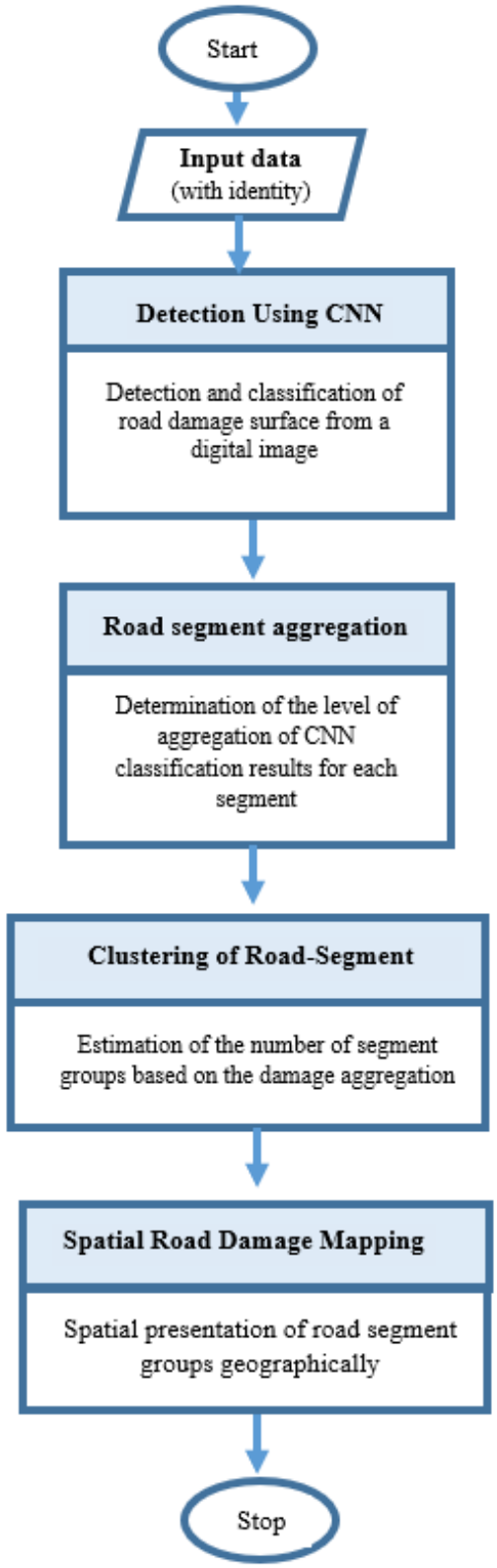


Figure 1. Stages of the Integrative Convolutional Neural Network–Assisted Clustering Framework (ICCF)

(iii) Clustering of road-segment

After feature aggregation, the ICCF scheme applies partition-based or hierarchical object clustering to group road segments with similar damage characteristics. The term CNN-assisted clustering refers to the use of features derived from CNNs and statistically aggregated as clustering inputs. Clustering is performed independently of and relies on conventional distance-based similarity measures from CNN-generated category aggregations.

This approach avoids the complexity and instability often associated with joint optimization or deep clustering models, while still leveraging the representational power of CNN-based image analysis. As a result, clustering results remain interpretable and reproducible, which is crucial for adoption by non-technical stakeholders.

(iv) Spatial road-damage mapping

The final stage of ICCF translates clustering results into valuable spatial information. Each cluster represents a group of road segments with similar damage profiles, rather than abstract statistical categories. In practical terms, the clustering output enables the identification of segments requiring urgent intervention, differentiation between preventive maintenance and major rehabilitation needs, and spatial visualization of road-condition patterns across the network.

By mapping clustered road segments to a geographic representation, the ICCF scheme supports maintenance prioritization, inspection scheduling, and resource allocation. This framework is particularly well-suited for local road authorities, as it provides a concise and easy-to-understand overview of the condition of the entire network without requiring extensive manual inspections or complex model interpretation.

3.2 Discussion on the scope and limitations of the Integrative Convolutional Neural Network–Assisted Clustering Framework scheme

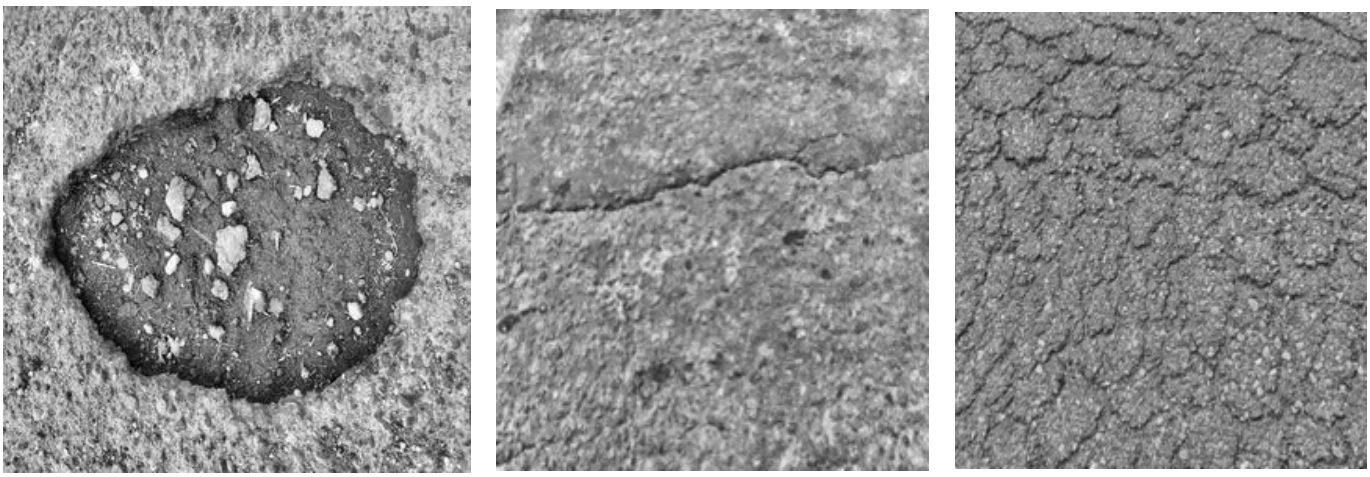
While ICCF enhances the usability of image-based road damage analysis, it does not replace detailed engineering assessments or cost-based evaluation models. The framework is designed for condition grouping and spatial mapping, and additional modules would be required to estimate repair costs or structural performance.

Nevertheless, ICCF demonstrates how established analytical tools can be systematically integrated to support scalable and data-driven road infrastructure management, especially in resource-constrained environments.

4. EXPERIMENTAL REALIZATION OF INTEGRATIVE CONVOLUTIONAL NEURAL NETWORK–ASSISTED CLUSTERING FRAMEWORK FOR SPATIAL-ROAD DAMAGE MAPPING

The ICCF scheme was implemented and evaluated using real-world road condition image data collected from 319 road segments in Sleman Regency, Yogyakarta. Each road segment contains multiple digital images captured under various lighting conditions and viewing angles, reflecting realistic operational constraints.

The three types of road damage image data used are as shown in Figure 2, namely (a) potholes, (b) corrugations, and (c) alligator cracks. The dataset utilized comprises 1335 images, which are categorized into three distinct classes.



(a) Pothole (b) Corrugation (c) Alligator crack

Figure 2. Example of road damage image

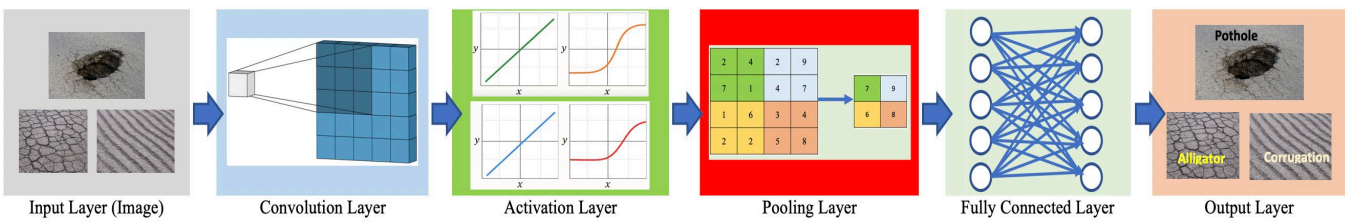


Figure 3. The outline of the Convolutional Neural Network (CNN) process

4.1 Convolutional Neural Network performance within the Integrative Convolutional Neural Network–Assisted Clustering Framework pipeline

In this study, the role of CNN is limited to robust feature extraction and damage categorization rather than architectural innovation. The CNN process for detecting and classifying road damage images is outlined in Figure 3.

The preprocessing stage of ICCF is a critical component for ensuring data consistency, quality, and computational efficiency prior to model training. Data augmentation is applied during training to enable the model to recognize and adapt to diverse image conditions, including suboptimal lighting and noisy inputs. This process generated approximately 9,345 image variations for the training set. Augmentation was applied exclusively to the training data; the validation and test sets were left unmodified to ensure an objective evaluation of model performance.

In the input layer, the data is interpreted as a 3-dimensional matrix (X), with the first and second dimensions representing the height (H) and width (W) of the image, while the third dimension represents the number of colors (C) in RGB or Grayscale format, as in formula (1):

$$X \in \mathbb{R}^{H \times W \times C} \tag{1}$$

Taking into account the variations in the five architectures used in this study, the input images were standardized using dimensions of 224 × 224 × 3. This size was chosen because it is lighter, allows for faster training, and has a relatively low risk of image deformation in the case of road textures. To control stability during the training process, the input data is normalized using Min-Max as in Eq. (2):

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{2}$$

The convolution layer in CNN functions to recognize patterns in images by detecting basic features such as edges, corners, and textures using filters/kernels. If there is an input like formula (1) and a kernel with the number of filters l like formula (3), then the convolution operation is like Eq. (4) with b as the bias and c as the input channel:

$$F \in \mathbb{R}^{l \times l \times C} \tag{3}$$

$$Y(i,j) = \sum_{m=1}^{l} \sum_{n=1}^{l} \sum_{c=1}^{C} X_{i+m-1,j+n-1,c} . F_{m,n,c} + b \tag{4}$$

where,

$Y(i,j)$: feature map value at position (i,j)

$X_{i+m-1,j+n-1,c}$: input value at position $(i+m, j+n)$

$F_{m,n,c}$: kernel value at position (m,n)

(m,n): index in kernel

The output dimension is affected by stride (s) and padding (p), such as in Eqs. (5) and (6):

$$H_{out} = \frac{H - l + 2p}{s} + 1 \tag{5}$$

$$W_{out} = \frac{W - l + 2p}{s} + 1 \tag{6}$$

Activation layers are used to help the model recognize complex patterns in input data, such as images. In this study, the Rectified Linear Unit (ReLU) activation function is applied because it is fast, simple, and does not cause gradient saturation. ReLU converts all negative values to 0, but for positive x values, the value returns to its original value, with the activation function as formula (7):

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{7}$$

The pooling layer chosen in this study is Max pooling because it can reduce the image size and maintain the dominant image area. For a 2 × 2 patch, max pooling is determined according to Eq. (8):

$$Y(i,j) = \max\{X(2i,2j), X(2i+1,2j), X(2i,2j+1), X(2i+1,2j+1)\} \tag{8}$$

The fully connected layer functions to calculate the final layer output in the CNN process, and changes the dimensions of the classifiable data linearly. In this case, the feature maps from the convolution and pooling processes are flattened into vectors in n-dimensional space according to formula (9):

$$X \in \mathbb{R}^{n} \tag{9}$$

The fully connected layer performs a linear transformation as in Eq. (10) and works like traditional classification:

$$z = Wx + b \tag{10}$$

where,

z: output before activation

W: matrix weight

x: input vector

The output layer is the final layer in a CNN architecture, producing the final model predictions and converting high-level feature representations (from previous layers) into a format that can be interpreted for specific tasks. In this study, the Softmax activation function is used to calculate the probability of an event, and the target class is determined based on the class with the highest probability. The Softmax function for the probability of the i-th class is shown in Eq. (11):

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{g} e^{z_j}}, i = j = 1, 2, \ldots, g \tag{11}$$

CNN prediction is the class with the highest probability as Eq. (12):

$$\hat{y} = \arg\max_{i} \; p_i \tag{12}$$

The five CNN architectures analyzed in this research include the second version of ResNet with 50 layers (ResNet50V2), the second version of ResNet with 152 layers (ResNet152V2), the VGGNet consisting of 16 layers (VGGNet-16), InceptionV3, and Xception. The reasons for using these five architectures in this study are as follows.

VGGNet-16 is a classic deep learning architecture that illustrates the effect of increased network depth on classification performance. It consists of 16 learnable layers — 13 convolutional layers with uniform 3 × 3 kernels and 3

fully connected layers — arranged in a simple, modular structure. Despite having approximately 138 million parameters, its architectural simplicity makes VGGNet-16 a widely used baseline and benchmark in deep learning research.

ResNet50V2 is a deep CNN that addresses the vanishing gradient problem through residual learning with skip connections. It comprises 50 layers organized into four residual stages (3-4-6-3) using bottleneck blocks for computational efficiency. These residual connections enable effective training of deep networks, contributing to strong performance in image classification and transfer learning tasks.

InceptionV3 employs inception modules with parallel convolutions of varying kernel sizes to capture multi-scale spatial features. The architecture incorporates factorized convolutions, auxiliary classifiers, and label smoothing to improve training efficiency while reducing parameter count. This design achieves a favorable balance between high accuracy and computational efficiency, making InceptionV3 suitable for practical applications.

Xception extends the Inception architecture by fully adopting depthwise separable convolutions, separating spatial and channel-wise feature extraction. This design simplifies the network structure while improving parameter efficiency and model expressiveness. Xception delivers strong classification performance with reduced computational cost, making it well-suited for resource-constrained environments.

ResNet152V2 is a very deep residual network with 152 layers that introduces pre-activation and refined bottleneck designs. These improvements enhance training stability and enable the extraction of more complex hierarchical features. As a result, ResNet152V2 achieves higher accuracy and better generalization, making it effective for challenging computer vision tasks.

Furthermore, each architecture is evaluated based on several metrics: accuracy, recall, precision, and F1-score. In this study, all evaluation metrics are reported in percentage (%). These metrics are measured by comparing the prediction results with the actual labels. For this reason, a confusion matrix for g classification is shown in Eq. (13):

$$CM = \begin{bmatrix} n_{11} & n_{12} & \cdots & n_{1g} \\ n_{21} & n_{21} & \cdots & n_{2g} \\ \vdots & \vdots & \ddots & \vdots \\ n_{g1} & n_{g2} & \cdots & n_{gg} \end{bmatrix} \tag{13}$$

where, n_{ij} denotes the number of objects that are actually class i, but are predicted to be class j.

Using the confusion matrix, the classification performance was evaluated in terms of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). TP refers to correctly predicted positive samples, TN refers to correctly predicted negative samples, FP refers to negative samples incorrectly classified as positive, and FN refers to positive samples incorrectly classified as negative. Based on these outcomes, Accuracy, Precision, Recall, and F1-score were used as evaluation metrics. Accuracy is calculated as shown in Eq. (14). Precision and Recall are defined in Eqs. (15) and (16), respectively. The F1-score, representing the harmonic mean of Precision and Recall, is given in Eq. (17).

$$\text{Accuracy} = \frac{\sum_{i=1}^{g} n_{ii}}{\sum_{i=1}^{g}\sum_{j=1}^{g} n_{ij}} \tag{14}$$

$$\text{Precision} = \frac{1}{g}\sum_{i=1}^{g} \frac{TP_i}{TP_i + FP_i} \tag{15}$$

$$\text{Recall} = \frac{1}{g}\sum_{i=1}^{g} \frac{TP_i}{TP_i + FN_i} \tag{16}$$

$$\text{F1} - \text{Score} = \frac{1}{g}\sum_{i=1}^{g} 2\frac{(Pr)_i (Rc)_i}{(Pr)_i + (Rc)_i} \tag{17}$$

Based on all the provisions above, this study produces a comparison of the performance of five architectures in terms of accuracy, precision, recall, and F1-Score values for three types of road damage, as shown in Figure 4. Overall, all models demonstrate strong and consistent performance, with metric values predominantly exceeding 90%, indicating good generalization on the validation dataset.

Among the evaluated architectures, ResNet152V2 achieved the highest performance across all four metrics, with values approaching 96%. This result is attributable to its deeper architecture combined with residual connections, which enable effective hierarchical feature extraction while mitigating vanishing gradient issues. The close alignment among accuracy, precision, recall, and F1-score further indicates balanced classification behavior without significant bias toward any specific class.

ResNet50V2 ranks second, showing performance only slightly lower than ResNet152V2. This suggests that residual learning remains highly effective even with a reduced depth, offering a favorable trade-off between model complexity and predictive accuracy. InceptionV3 and Xception exhibit comparable performance, with validation metrics in the mid-94% range. Their architectural designs, which emphasize multi-scale feature extraction and depthwise separable convolutions, respectively, contribute to robust classification results, although marginally lower than the deeper residual networks. In contrast, VGGNet demonstrates the lowest validation performance, with metrics around 91–92%. The absence of residual or advanced feature fusion mechanisms likely limits its ability to capture complex patterns, particularly when compared to more modern architectures.

Furthermore, Figure 5 illustrates the comparative performance of the evaluated CNN architectures on the testing dataset using accuracy, precision, recall, and F1-score. In contrast to the validation results, a general performance decline is observed across all models, which is expected when models are evaluated on completely unseen data and provides a more realistic assessment of generalization capability.

ResNet50V2 achieves the best overall performance on the test data, with all evaluation metrics exceeding 92%. This result indicates that ResNet50V2 offers the most robust generalization, likely due to its balanced architectural depth that effectively captures discriminative features while mitigating overfitting. ResNet152V2 follows closely, maintaining strong performance slightly above 91%. Although deeper networks perform well during validation, the marginal decrease in testing data suggests that increased depth does not necessarily translate into superior generalization under limited or diverse testing samples. InceptionV3 and Xception show moderate performance, with metric values in the range of approximately 86–90%. Their consistent but lower results indicate stable learning behavior, though with reduced

discriminative power compared to residual-based architectures. In contrast, VGG exhibits the lowest testing performance, with all metrics around 75–76%. The absence of residual connections and the relatively shallow feature representation likely limit its ability to generalize to unseen data.

Although several CNN architectures achieved comparable validation metrics, a closer examination reveals differences in generalization behavior across models. The identical validation accuracy observed in some cases can be attributed to the use of a balanced dataset and macro-averaged performance measures, which may mask subtle variations between classes when evaluated on validation subsets. In contrast, testing results exhibit larger variance across models, indicating differences in robustness and generalization capacity. Deeper architectures, such as ResNet152V2, demonstrate a tendency toward overfitting, as reflected by a noticeable performance drop when evaluated on unseen data. This behavior is likely due to the limited diversity of training samples relative to the model complexity, despite data augmentation strategies.

From a practical standpoint, model selection in the proposed ICCF framework prioritizes stability and generalization over marginal gains in validation accuracy. While deeper models offer higher representational capacity, lighter architectures such as ResNet50V2 provide a more favorable balance between accuracy, computational efficiency, and robustness. Consequently, ResNet50V2 is selected for integration within the ICCF framework, particularly for deployment in resource-constrained environments where computational efficiency and reliability are critical considerations.

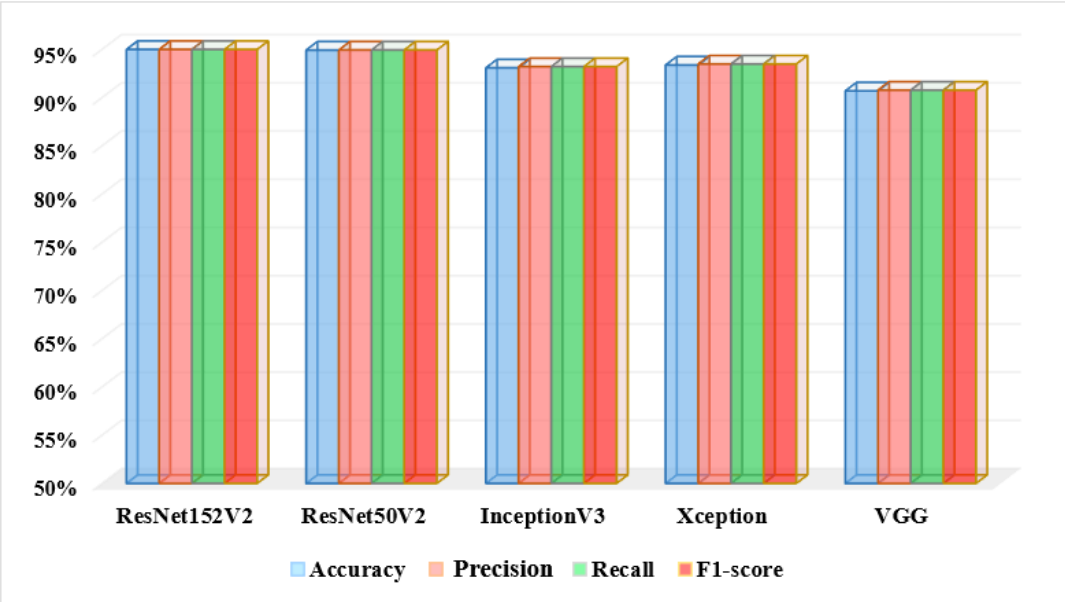


Figure 4. Architecture performance for validation data

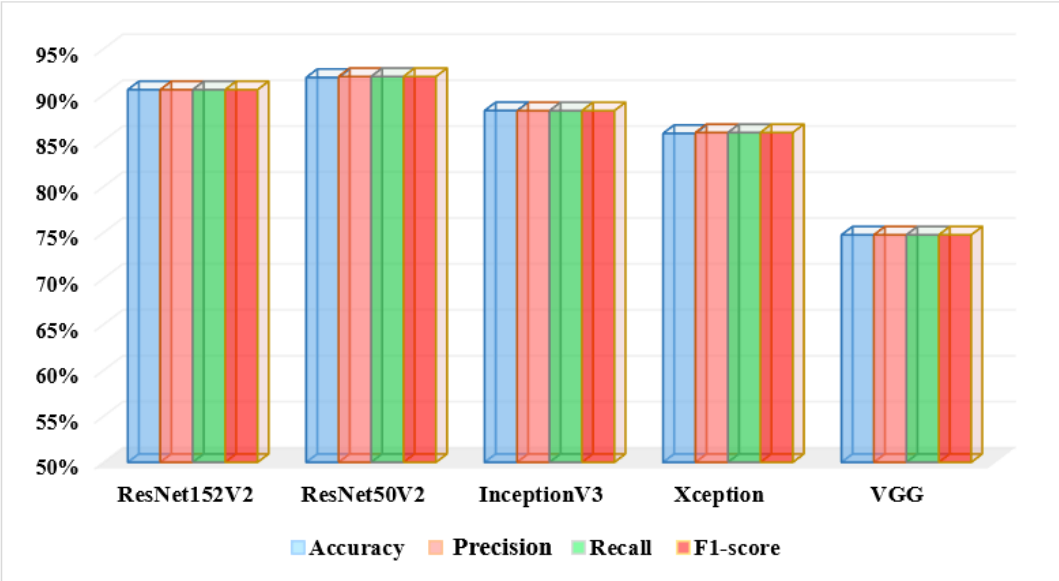


Figure 5. Architecture performance for testing data

4.2 Segment-level aggregation results

The different data, which has never been used before in the CNN process in the first stage of ICCF, was taken from 108 road segments in Sleman Regency, with a total of 2,319 images. The results of CNN detection and classification using the ResNet152V2 architecture were then aggregated for each road segment. For example, if n_s denotes the total number of damages on the s-th road segment, and f_d denotes the number of damages for the d-th damage type on the s-th segment, then the proportion of the d-th damage in the s-th segment is calculated according to Eq. (18):

$$P_{s_d} = \frac{f_d}{n_s} \tag{18}$$

This study only uses three types of road damage as a grouping, so that a three-dimensional plot can be presented using the f_d value for each segment as in Figure 6. The aggregation results in Figure 6 will later be used as the basis for input for road segment mapping.

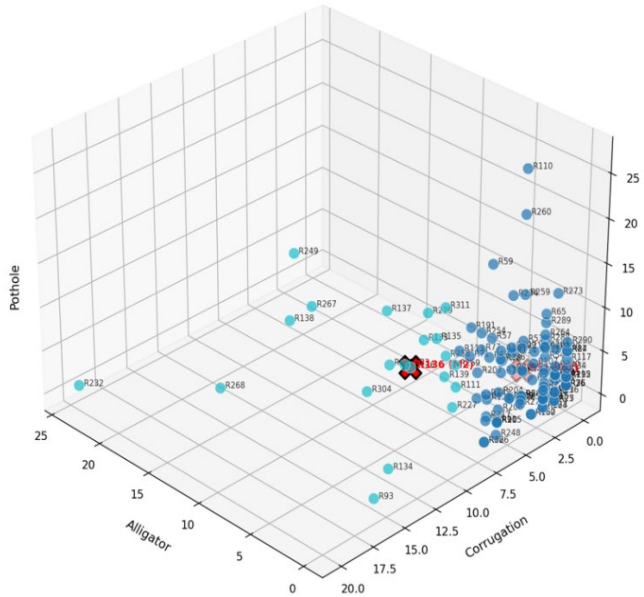


Figure 6. Profile of the aggregation level of road damage

4.3 Clustering outcomes and determining the number of groups

4.3.1 Justification of the selection of the partition method for Integrative Convolutional Neural Network–Assisted Clustering Framework for the spatial road-segment

The three main stages in cluster analysis are pre-processing, algorithm selection, and validation of clustering results. In this study, pre-processing was carried out by standardizing the data using Eq. (2). K-means, block-based k-medoids (Block-KM), modified fast k-medoids (MFast-KM), or other methods can basically be applied in the third stage of the ICCF scheme. This study uses a partition-based method to cluster objects (road segments) based on the aggregation of damage levels. This technique was chosen due to its efficiency and ability to work directly with high-dimensional data.

To select the most appropriate partitioning method, three candidate methods were evaluated using synthetic data with characteristics resembling the CNN aggregation results. The use of synthetic data in this context is intended solely as a supplementary robustness assessment of the clustering mechanism, not as a proxy for real-world road imagery features. The artificial data were arranged into three cluster scenarios, and each model was replicated fifty times. The method evaluation was based on the Adjusted Rand Index, Rand Index, Jaccard Coefficient, Hubert Statistic, and Folkes-Mallow values. The Manhattan distance between two objects (say A and B) that have the variable p is as shown in Eq. (19):

$$d_{AB} = \sum_{v=1}^{p} |x_{Av} - x_{Bv}| \tag{19}$$

Three groups of 100 objects each are observed in a bivariate normal distribution with cluster centers at (150, 150), (150, 250), and (250, 175), with each having the same standard deviation of twenty. A group with an elongated pattern that resembles a road section, where each section is cut into segments measuring, for example, 20 square meters, is shown in Figure 7.

Nine groups of 100 each are observed on two normally distributed variables. Three horizontal groups for three positions from the bottom, middle, and top, each has the same center for the first variable, namely $(150, 200\sqrt{2}, 300\sqrt{2})$. While the center of the second variable has the same value for the three horizontal groups, at the bottom position is zero, the same in the middle position is $200\sqrt{2}$, and the same in the top position is $300\sqrt{2}$. Three vertical groups on the left and right sides are applied a standard deviation of 20, while for three vertical groups in the middle position, a standard deviation of 25 is applied.

Figure 8 shows that in general, the three clustering methods, namely k-means, Block-KM, and MFast-KM, achieve excellent performance in artificial data scenarios with a relatively small number of groups (3 and 5 groups), as indicated by Rand Index, Adjusted Rand Index, Hubert's Statistics, Jaccard Coefficient, and Fowlkes-Mallows Index values approaching 1.0. This indicates that the cluster structure in the data in the first two scenarios is relatively easy for these methods to learn.

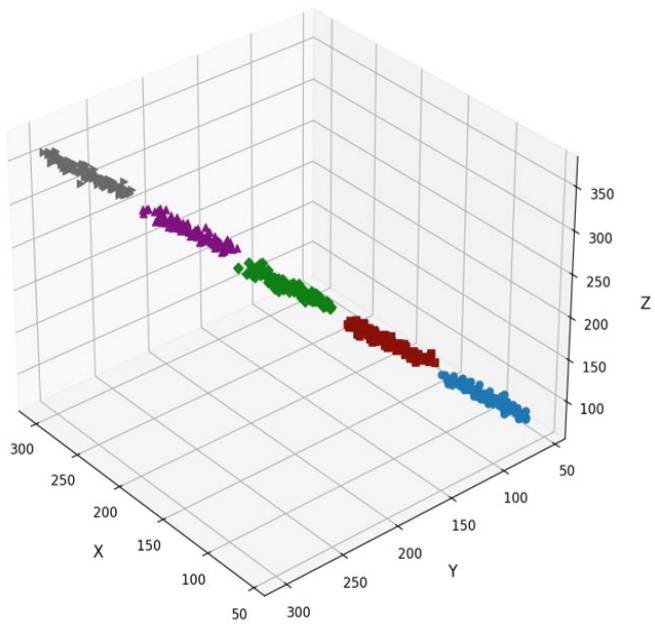
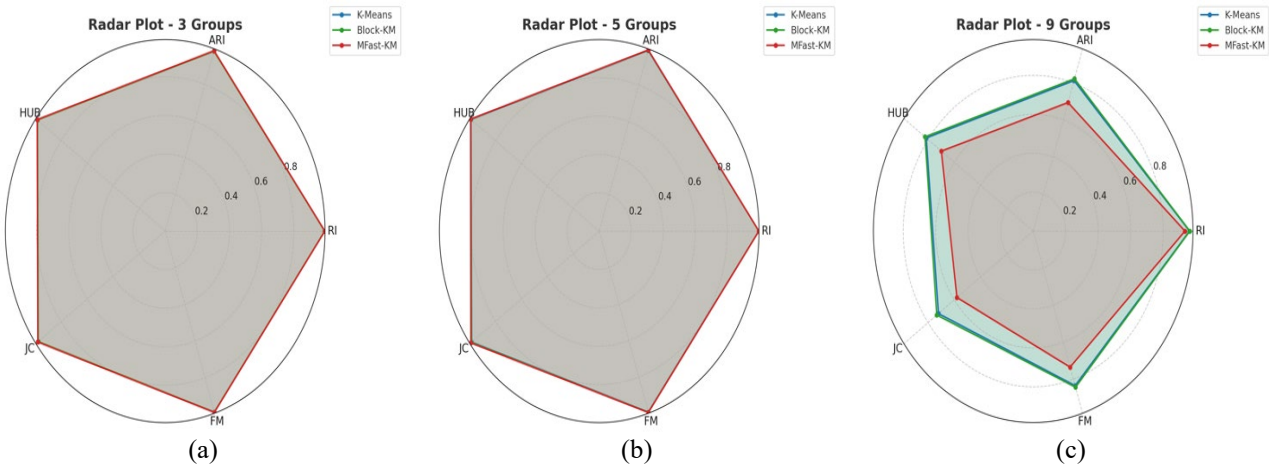


Figure 7. Artificial data resembling road segments



(a) (b) (c)

Figure 8. Clustering performance

In the first scenario (3 groups, Figure 8(a)), MFast-KM demonstrated the highest scores for almost all metrics, followed by k-means and Block-KM with very small differences (≤ 0.002). Meanwhile, in the second scenario (5 groups, Figure 8(b)), overall performance improved across all methods, with MFast-KM again achieving good results. This indicates that the model can maintain stability and robustness when the number of clusters increases moderately.

However, significant performance differences began to emerge in the third scenario (9 groups, Figure 8(c)). ARI and JC values consistently decreased across all methods, indicating the greater modeling challenges of mapping increasingly complex clusters. In this scenario, Block-KM performed relatively well (ARI = 0.8225; JC = 0.7335; FM = 0.8432), followed by k-means, while MFast-KM experienced the sharpest decline (ARI = 0.6941; JC = 0.5802; FM = 0.7347).

Overall, these results confirm that all methods are highly effective for low to medium cluster sizes (3–5 clusters). Block-KM can adapt to increasingly complex cluster structures (9 clusters). MFast-KM excels on simpler data, but experiences a significant performance decline as the number of clusters increases. Based on these results, the Block-KM method was applied in this study. Spatial analysis was used to map the damage locations on each road segment, and this spatial information allows for more accurate identification of damage-prone points. Incorporating spatial constraints or graph-based clustering methods is an important direction for future research.

4.3.2 Determining the number of clusters

To determine the optimal number of clusters, this study employs the Silhouette coefficient, gap statistic, and Calinski–Harabasz index. These indices are widely used internal validation measures that evaluate clustering quality based on compactness, separation, and relative dispersion, without requiring external ground truth labels. The Silhouette coefficient assesses the consistency of data assignment within clusters, while the gap statistic compares within-cluster variation to a reference null distribution. The Calinski–Harabasz index complements these measures by quantifying the ratio between inter-cluster separation and intra-cluster cohesion, providing a balanced criterion for selecting the number of clusters.

For each data point (road segment) i, the Silhouette coefficient is defined as Eq. (20). The overall Silhouette score is obtained by averaging $s(i)$ over all data points.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{20}$$

where, $a(i)$ denotes the average distance between data point i and all other points within the same cluster (intra-cluster distance). $b(i)$ represents the minimum average distance between data point i and all points in the nearest neighboring cluster (inter-cluster distance).

The variance ratio criterion (VRC) is constructed by Calinski-Harabasz, as in Eq. (21) [30]:

$$VRC = \frac{\left(\bar{d}^2 + A_k(n-k)/(k-1)\right)}{\left(\bar{d}^2 - A_k\right)} \tag{21}$$

where,

$$A_k = \frac{1}{n-k}\sum_{g=1}^{k}\left(n_g - 1\right)\left(\bar{d}^2 - \bar{d}_g^2\right) \tag{22}$$

The gap statistic evaluates the clustering structure by comparing the within-cluster dispersion (W_k) of the observed data with that of a reference null distribution, such as in Eq. (23):

$$\text{Gap}(k) = \frac{1}{B}\sum_{b=1}^{B}\log\left(W_k^{*(b)}\right) - \log\left(W_k\right) \tag{23}$$

where, W_k is the within-cluster dispersion for k clusters computed from the observed data. $W_k^{*(b)}$ denotes the within-cluster dispersion obtained from the b-th reference dataset generated under a null model. B is the total number of reference datasets.

All criteria produced the same conclusion, namely that the ideal group size to be formed is two groups, as shown in Figure 9. In the Silhouette Score, the highest value is seen at k = 2 with a value close to 0.46, which indicates that in this condition, the cluster structure has the most optimal level of compactness within the cluster and the level of separation between clusters compared to other numbers of clusters. When the number of clusters is increased from k = 3 to k = 8, the Silhouette value tends to decrease or fluctuate at a lower number, thus indicating a decrease in the quality of data separation into more heterogeneous groups.

Meanwhile, the Calinski-Harabasz criterion also supports

these results, where the highest score is at k = 2 with a value of more than 52. After the number of clusters increases, the value of this index is unable to exceed the value at k = 2, which means that the level of separation between clusters and the strength of the internal structure of the clusters remains best at the division of two groups.

The third evaluation, namely the gap statistic, also shows a consistent pattern, where the highest gap statistic value is found at k = 2, while at a larger number of clusters, there is a decrease or instability in the value of this metric, accompanied by a larger deviation. The combination of these three indices provides strong evidence that the road segment datasets in this study form two main groups with the most clearly separated structures.

From a practical and managerial perspective, classifying road segments into two condition groups provides a meaningful and operationally relevant framework for road authorities. In routine infrastructure management, decision-making is often constrained by limited budgets, time, and personnel. Road agencies therefore tend to prioritize actions based on broad condition categories rather than fine-grained technical distinctions that are difficult to operationalize.

In this context, clustering road segments into two characteristic groups aligns well with common maintenance planning practices, where roads are typically classified into (i) segments requiring immediate or prioritized intervention and (ii) segments that remain serviceable and can be managed through routine monitoring or preventive maintenance. This binary grouping supports rapid screening of the road network and enables decision-makers to allocate resources more efficiently without the need for complex multi-class assessments.

The ICCF scheme facilitates this process by transforming image-level damage detections into segment-level profiles that reflect the overall severity and distribution of surface defects. By grouping these profiles into two clusters, the resulting road-condition map provides a clear and interpretable representation of network-level conditions. Such representation is more actionable than descriptive tables alone, as it highlights spatial patterns of deterioration and supports strategic prioritization at the network scale.

Therefore, the use of two clusters is not intended to oversimplify road conditions, but rather to reflect the realities of infrastructure management workflows. The resulting classification also serves as a decision-support tool that bridges detailed image-based analysis and high-level maintenance planning, making it particularly suitable for use by local road agencies and public works departments operating in resource-constrained environments.

To emphasize the cluster structure in a two-dimensional form that is easy to present, the Principal Component Analysis (PCA) of the biplot is shown in Figure 10.

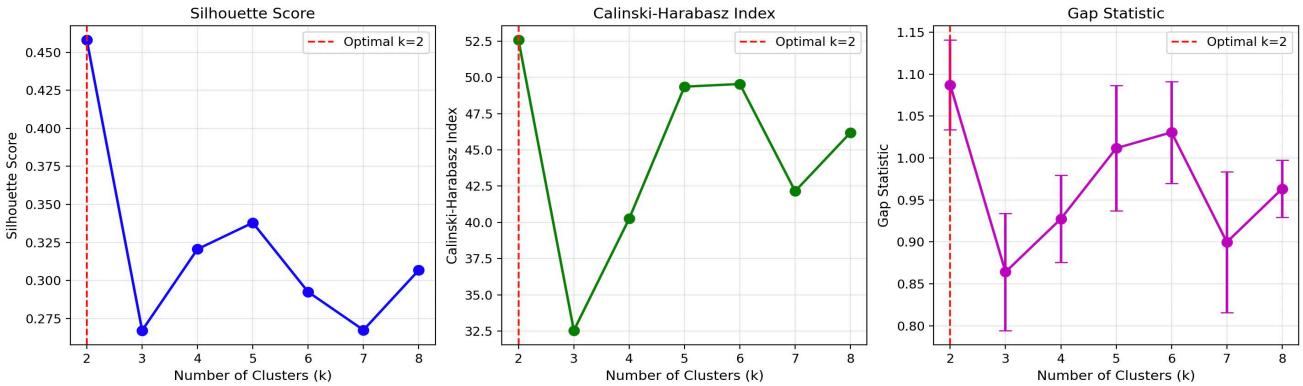


Figure 9. Cluster evaluation

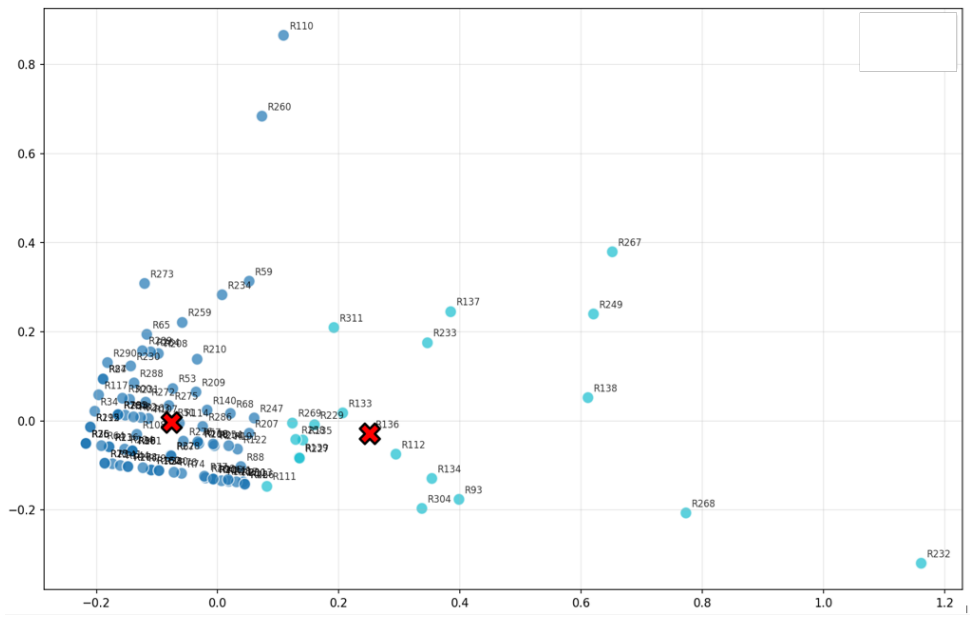


Figure 10. Biplot of Principle Component Analysis (PCA) for clustering

Figure 10 shows the results of applying the k-medoids method to group data into two clusters and reduce its dimensionality with PCA into a two-dimensional space (PC1 and PC2) that explains 82.3% of the data variation.

Visually, the two clusters (Cluster 1 and Cluster 2) appear quite separate, although there are still objects with close positions in the transition area between clusters. The large red dots with a cross (X) in each cluster are medoids, namely, reference objects that are the center of the group according to the characteristics of the original data (not the average value as in k-means). Cluster 1 dominates most of the area around PC1, negative to near zero, with a fairly dense distribution and a pattern that tends to form a sideways wave.

Meanwhile, Cluster 2 is located on the right side of the plot with a higher PC1 value, indicating that the main characteristics of objects in this group have a more positive PC1 score so that they are significantly different from the objects in Cluster 1. The distribution of points in both clusters shows a clearer group structure in the first dimension (PC1) compared to PC2, which indicates that the main character differences between objects are largely explained by the first principal component. Overall, the clustering results indicate the presence of two relatively consistent groups, visually and medoidally located at representative positions of each cluster, thus providing an interpretation that the separation of these two groups is quite good in explaining the pattern of data variability.

The three-dimensional plot (Figure 6) illustrates the results of clustering using the K-Medoids method based on the three original variables: the number of potholes, corrugation, and alligator damage types. This visualization also shows the distribution of road objects in three-dimensional space without transformation or dimensionality reduction, thus providing a more direct picture of the data patterns. The graph shows two clusters distinguished by different colors: the first cluster mostly includes road sections with low to moderate damage frequencies, especially in the corrugation and pothole dimensions, while the second cluster generally has higher corrugation values and is located in a denser and closer area, indicating a stronger homogeneity of characteristics. Large dots with red crosses mark the positions of the medoids of each cluster, namely representative objects that best describe the pattern of each group. The separation of the two clusters is quite visible in the combination of the corrugation and pothole dimensions, although the points in cluster 1 are more widely distributed and tend to reflect a greater variety of damage conditions.

Overall, this 3D plot confirms the presence of two main characteristics of road damage conditions: a group of roads that predominantly experience light–moderate damage and a group of roads that tend to be more severely corrugation-damaged, and these results support the visual interpretation that the cluster structure is naturally formed from the data.

4.4 Spatial interpretation and application for road maintenance

Based on the distribution map of road damage clusters (Figure 11), a spatial pattern is visible, indicating that sections included in Cluster 2 (marked with a thick red line) tend to be concentrated in the northern and northwest parts of Sleman Regency, namely the areas closer to Mount Merapi. This area is the main distribution route for volcanic materials, such as sand and stone, which are intensively transported by heavy trucks. The excessive load and frequent movement of these transport vehicles contribute to creating repeated stress and accelerating the degradation of the road pavement layer. Meanwhile, the road sections in Cluster 1 (thick black line) are spread relatively evenly across the central to southern regions of Sleman Regency, Yogyakarta, Indonesia, on road sections that are not the main routes for mining truck traffic, so the level of damage is lower.

This finding indicates that economic activities based on the utilization of Merapi's resources have a significant influence on the level of damage to road infrastructure, making mining logistics routes a top priority in planning for maintenance and improvement of pavement quality by the local government.

The interpretation of the resulting clusters is intended to support network-level screening and prioritization rather than to replace detailed engineering assessments. While individual metrics such as repair cost or severity indices are not explicitly computed in this study due to data availability constraints, the clustering results provide an intermediate decision-support layer that aggregates image-level damage information into interpretable road-segment profiles.

Compared with a simple summary table, the clustering-based mapping offers added value by revealing patterns of similarity and contrast among road segments that are not immediately apparent from descriptive statistics alone. Road segments within the same cluster share comparable damage composition and distribution characteristics, enabling road authorities to identify groups of segments likely to require similar levels of intervention or monitoring. The figures presented in this study collectively illustrate the stages and rationale for component selection within ICCF, as well as the end-to-end behavior of the proposed ICCF implementation, from image-level damage detection to network-level road condition mapping.

From an operational perspective, the clustered road-condition map facilitates rapid prioritization by highlighting segments that exhibit consistently higher concentrations of surface damage relative to the rest of the network. This information can be used as a preliminary filter to guide more detailed inspections, budgeting analyses, or cost-based evaluations conducted by road agencies. As such, the ICCF framework complements, rather than replaces, traditional engineering metrics by providing a scalable and interpretable means of translating image-based detections into actionable spatial insights.

The experimental results confirm that ICCF effectively integrates CNN-based image analysis with a clustering framework without altering the underlying algorithms. By aggregating predictions at the segment level, ICCF reduces image-level uncertainty and yields more stable and interpretable groupings of road conditions. This structured integration allows heterogeneous visual information to be consolidated into consistent representations that are meaningful at the road-segment scale.

These outcomes indicate that the main contribution of ICCF lies in its operational integration and application-oriented design rather than in methodological novelty. The framework is particularly suited to supporting spatial decision-making in road maintenance planning, where reliability, interpretability, and ease of implementation are more critical than introducing new learning algorithms.

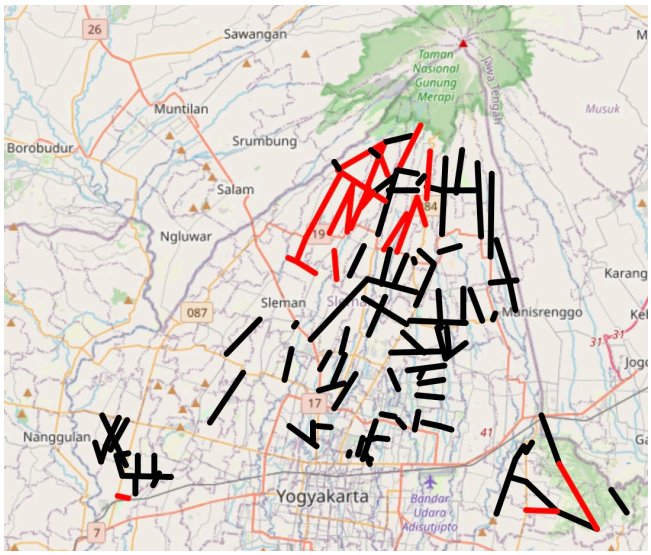


Figure 11. Spatial pattern of road damage condition

5. CONCLUSION

This study presented ICCF, an integrative analytical pipeline for spatial road damage mapping using digital image data. The principal contribution of this work is the systematic integration of CNN-based image-level damage detection, structured statistical aggregation, and partition-based clustering into a unified framework that supports network-level road-condition analysis.

Experimental results confirm that the proposed framework is capable of transforming large volumes of road imagery into interpretable segment-level damage profiles. Among the evaluated CNN architectures, ResNet50V2 provided the most favourable balance between accuracy, generalization, and computational efficiency, making it the most suitable candidate for deployment within the ICCF pipeline. The subsequent clustering analysis enables the grouping of road segments into two characteristic condition categories that are meaningful from a maintenance planning perspective, supporting rapid network-level screening and prioritization.

From a practical standpoint, the ICCF framework should be viewed as a decision-support tool rather than a replacement for detailed engineering assessment or cost-based optimization models. The clustering-based road-condition maps provide added value by revealing spatial patterns and similarities among road segments that are not readily apparent from descriptive summary tables alone. These outputs can assist road authorities in identifying segments that warrant further inspection, detailed evaluation, or prioritized intervention.

The figures presented in this study collectively illustrate the stages and rationale for component selection within ICCF, as well as the end-to-end behaviour of the proposed ICCF implementation, from image-level damage detection to network-level road condition mapping. Performance comparison figures highlight the relative stability and generalizability of different CNN architectures, supporting the selection of ResNet50V2 as a balanced backbone for integration within the framework. Comparison of group count criteria validates the size of road segment damage clusters.

From an applied perspective, the spatial representation provided by the clustering maps offers insights beyond tabulated summaries. Together, these visual and numerical elements reinforce ICCF's role as a decision-support framework that bridges detailed image analysis and high-level infrastructure management considerations.

Three methodological limitations should be acknowledged. First, the clustering stage operates on aggregated statistical representations rather than learned latent embeddings; it therefore does not constitute deep embedded clustering in the strict theoretical sense. Second, spatial adjacency and network topology among road segments are not explicitly modelled, which may limit the ability to capture spatial autocorrelation effects. Third, actionable engineering metrics — such as repair cost and severity-weighted indices — are not directly computed due to data availability constraints.

Future research may address these limitations by incorporating spatial constraints or graph-based clustering, integrating cost- or severity-weighted damage indicators, and exploring tighter coupling between feature learning and clustering in a semi-supervised setting. Despite these limitations, the present study demonstrates that an integrative, application-oriented framework can effectively bridge image-based damage detection and spatial road-condition mapping for infrastructure management in resource-constrained environments.

ACKNOWLEDGMENT

The author would like to express gratitude to the Ministry of Higher Education, Science and Technology of the Republic of Indonesia for the financial support that has been provided, as stated in Decree Number 126/C3/DT.05.00/PL/2025; 0498.01/LL5-INT/AL.04/2025; and the research contract letter in the framework of the Implementation of the Fundamental Research Program–Regular Research Scheme Number: 029/ST-DirDPPM/70/DPPM/PFR-KEMDIKTI SAINTEK/VI/2025.

REFERENCES

[1] Bridgelall, R. (2023). Driving standardization in infrastructure monitoring: A role for connected vehicles. Vehicles, 5(4): 1878-1891. https://doi.org/10.3390/vehicles5040101

[2] Kuncoro, E., Wurarah, R.N, Erari, I.E. (2024). The impact of road infrastructure development on ecosystems and communities. Social, Ecological, Economy for Sustainable Development Goals Journal, 1(2): 78-90. https://doi.org/10.61511/seesdgj.v1i2.2024.336

[3] Sholevar, N., Golroo, A., Esfahani, S.R. (2022). Machine learning techniques for pavement condition evaluation. Automation in Construction, 136: 104190. https://doi.org/10.1016/j.autcon.2022.104190

[4] Islam, M.S., Ibrahim, A.M., Haque, K.E., Bakhuraisa, K.A., Ali, U., Skitmore, M. (2024). Advancement in the automation of paved roadways performance patrolling: A review. Measurement, 232: 114734. https://doi.org/10.1016/j.measurement.2024.114734

[5] Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., Omata, H. (2018). Road damage detection and classification using deep neural networks with smartphone images. Computer-Aided Civil and Infrastructure Engineering, 33(12): 1127-1141. https://doi.org/10.1111/mice.12387

[6] Triardana, Y., Sasmita, B. Hadi, F. (2021). Road damage

identification using deep learning method convolution neural networks model. Jurnal Geodesi Undip, 10(3): 33-40. https://doi.org/10.14710/jgundip.2021.31159

[7] Que, Y., Dai, Y., Ji, X., Leung, A.K., Chen, Z., Jiang, Z., Tang, Y. (2023). Automatic classification of asphalt pavement cracks using a novel integrated generative adversarial networks and improved VGG Model. Engineering Structures, 277(15): 115406. https://doi.org/10.1016/j.engstruct.2022.115406

[8] Lamichhane, B.R., Srijuntongsiri, G., Horanont, T. (2025). CNN based 2D object detection techniques: A review. Frontiers in Computer Science, 7: 1437664. https://doi.org/10.3389/fcomp.2025.1437664

[9] Rajput, P., Chaturvedi, M., Patel, V. (2022). Road condition monitoring using unsupervised learning based bus trajectory processing. Multimodal Transportation, 1(4): 100041. https://doi.org/10.1016/j.multra.2022.100041

[10] Khorolska, K., Lazorenko, V., Bebeshko, B., Desiatko, A., Kharchenko, O., Yaremych, V. (2022). Usage of clustering in decision support system. Lecture Notes in Networks and Systems, 213: 615-629. https://doi.org/10.1007/978-981-16-2422-3_49

[11] Schubert, E., Rousseeuw, P.J. (2021). Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms. Information Systems, 101: 101804. https://doi.org/10.1016/j.is.2021.101804

[12] Kariyam, Abdurakhman, Effendie, A.R. (2025). Comparison of several clustering methods in classifying countries based on the environmental performance index. AIP Conference Proceedings, 3248: 040001. https://doi.org/10.1063/5.0236690

[13] Xuansen, H., Fan, H., Yueping, F., Lingming, J., Runzong, L., Allam, M. (2023). An effective clustering scheme for high-dimensional data. Multimedia Tools and Applications, 83(15): 1-45. https://doi.org/10.1007/s11042-023-17129-4

[14] Li, G., Xin, Y., Shen, D., Wang, B., Deng, Y., Zhang, S. (2023). Automatic road crack detection and analysis system based on deep feature fusion and edge structure extraction. International Journal of Pavement Engineering, 24(1). https://doi.org/10.1080/10298436.2023.2246096

[15] Bayane, I., Leander, J., Karoumi, R. (2024). An unsupervised machine learning approach for real-time damage detection in bridges. Engineering Structures, 308: 117971. https://doi.org/10.1016/j.engstruct.2024.117971

[16] Zhenglong, L., Hao, Z., Zhu, Y., Lu, C. (2025). A Review on automated detection and identification algorithms for highway pavement distress. Applied Sciences, 15(11): 6112. https://doi.org/10.3390/app15116112

[17] Trojanowski, P., Husár, J., Hrehová, S., Adamczak, M., Kolinski, A. (2024). Cluster analysis as a basis for the development of an application assessing the reliability of transport infrastructure. Mobile Networks and Applications, 29: 981-990. https://doi.org/10.1007/s11036-024-02328-6

[18] Kariyam, Abdurakman, Effendie, A.R. (2024). Modified fast K-Medoids to guarantee no empty group. AIP Conference Proceedings (Indexed Scopus Proceeding), 3201: 060009. https://doi.org/10.1063/5.0230965

[19] Pawar, K., Attar, V. (2022). Deep learning based detection and localization of road accidents from traffic surveillance videos. ICT Express, 8(3): 379-387. https://doi.org/10.1016/j.icte.2021.11.004

[20] Pendyala, M., Ananth, P., Natarajan, P., Somasundaram, K., Rajkumar, E.R., Ravichandran, K.S., Balasubramanian, V., Gandomi, A.H. (2024). An analysis of causative factors for road accidents using partition around medoids and hierarchical clustering techniques. Engineering Reports, 6(6): e12793. https://doi.org/10.1002/eng2.12793

[21] Meftah, I., Hu, J., Asham, M.A., Meftah, A., Zhen, L., Wu, R. (2024). Visual detection of road cracks for autonomous vehicles based on deep learning. Sensors, 24: 1647. https://doi.org/10.3390/s24051647

[22] Sivanarayana, G.V., Kumar, K.N., Srinivas, Y., Raj Kumar, G.V.S. (2021). Review on the methodologies for image segmentation based on CNN. Lecture Notes in Networks and Systems, 134: 165-175. https://doi.org/10.1007/978-981-15-5397-4_18

[23] Li, P. Pei, Y., Li, J. (2023). A comprehensive survey on design and application of autoencoder in deep learning. Applied Soft Computing, 138: 110176. https://doi.org/10.1016/j.asoc.2023.110176

[24] Lu, W., Qian, M., Xia, Y., Lu, Y., Shen, J., Fu, Q., Lu, Y. (2024). Crack_PSTU: Crack detection based on the U-Net framework combined with Swin Transformer. Structures, 62: 106241. https://doi.org/10.1016/j.istruc.2024.106241

[25] Xie, J., Girshick, R., Farhadi, A. (2015). Unsupervised deep embedding for clustering analysis. arXiv: 1511.06335. https://doi.org/10.48550/arXiv.1511.06335

[26] Wang, L., Zhang, M., Gao, X., Shi, W. (2024). Advances and challenges in deep learning-based change detection for remote sensing images: A review through various learning paradigms. Remote Sensing, 16(5): 804. https://doi.org/10.3390/rs16050804

[27] Diallo, B., Hu, J., Li, T., KHan, G.A., Liang, X., Zhao, Y. (2021). Deep embedding clustering based on contractive autoencoder. Neurocomputing, 433: 96-107. https://doi.org/10.1016/j.neucom.2020.12.094

[28] Pally, R.J., Samadi, S. (2022). Application of image processing and convolutional neural networks for flood image classification and semantic segmentation. Environmental Modelling & Software, 148: 105285. https://doi.org/10.1016/j.envsoft.2021.105285

[29] Alabidi, S.A., Al-Zubaidi, E.A. (2024). Satellite image classification using unsupervised machine learning. Al-Furat Journal of Innovations in Electronics and Computer Engineering, 3(2): 276-298. https://doi.org/10.46649/fjiece.v3.2.19a.28.5.2024

[30] Kariyam, Abdurakhman, Effendie, A.R. (2023). A Medoid-based deviation ratio index to determine the number of clusters in a datasets. MethodsX, 10: 102084. https://doi.org/10.1016/j.mex.2023.102084