Alexandria Engineering Journal 92 (2024) 310–320

Contents lists available at ScienceDirect

Alexandria Engineering Journal

journal homepage: www.elsevier.com/locate/aej

Original article

An investigation into real-time surface crack classification and measurement for structural health monitoring using transfer learning convolutional neural networks and Otsu method

Mazleenda Mazni [a,b,*], Abdul Rashid Husain [a], Mohd Ibrahim Shapiai [c], Izni Syahrizal Ibrahim [d], Devi Willieam Anggara [e], Riyadh Zulkifli [a]

[a] *School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia*
[b] *Faculty of Mechanical Engineering, Universiti Teknologi MARA Cawangan Johor, Kampus Pasir Gudang, Jalan Purnama, Masai, Malaysia*
[c] *Centre for Artificial Intelligence and Robotics, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia*
[d] *Forensic Engineering Centre, Institute for Smart Infrastructure and Innovative Construction, Faculty of Civil Engineering, Universiti Teknologi Malaysia, Johor, Malaysia*
[e] *School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Malaysia*

ARTICLE INFO

Keywords:
Crack classification
Crack segmentation
Structural Health Monitoring
Transfer Learning
Otsu Method
Convolution neural networks

ABSTRACT

This study introduces a pioneering system for real-time classification and measurement of concrete surface cracks, a crucial aspect of Structural Health Monitoring (SHM). We harness the power of transfer learning (TL) in Convolutional Neural Networks (CNNs), including renowned models such as MobileNetV2, EfficientNetV2, InceptionV3, and ResNet50. Notably, our model excels, particularly with TL MobileNetV2, achieving remarkable results – a 99.87% accuracy rate, 99.74% recall, 100% precision, and an impressive 99.87% F1-score. Incorporating the Otsu method for image segmentation, our system accurately assesses individual crack sizes. To refine measurements, Euclidean distance calculations and a 'pixel per inch' technique, accounting for video resolution, ensure millimeter-level width estimations. Precision is validated through manual experiments using a vernier caliper, specifically the Mitutoyo Absolute Digital Caliper. This tool ensures high accuracy, with an error margin of ±0.2 mm to ±0.3 mm, making it efficient for detailed measurements. Despite these promising outcomes, it is crucial to acknowledge inherent limitations. These include dependence on image quality, challenges in generalization, sensitivity to training data, assumption of linear crack width calculation, resolution dependency, and other factors. These limitations underscore the need for further refinement in our proposed classification model and measurement technique. This research represents a significant advancement in the SHM field, catering to early detection and timely maintenance requirements essential for infrastructure safety and longevity. As the field increasingly prioritizes rapid detection, our model presents a versatile solution that enhances the potential of Structural Health Monitoring.

1. Introduction

In the domain of civil engineering, Structural Health Monitoring (SHM) stands on the cusp of revolutionizing the industry by offering cutting-edge capabilities [1]. This method encompasses the consistent monitoring of a structure or mechanical system by collecting measurements at regular intervals, isolating characteristics that are sensitive to potential damage from these measurements, and utilizing statistical analysis to evaluate the system's current health status [2].Substantial investments are made in procuring a variety of instruments and resources to identify flaws in vital infrastructure like roads, bridges, buildings, and bodies of water, amounting to significant financial outlays [3]. Concrete cracks are a widespread issue in solid structures like building walls, roofs, bridges, and tunnels. Detecting these cracks early on is of utmost importance since they serve as indicators of the structural integrity of concrete infrastructure.

Cracks serve as early warning signs of aging, deterioration, or potential internal structural flaws within solid surfaces [4]. Internal structural damages can substantially weaken the integrity and durability of concrete structures, rendering them vulnerable to potential structural failure, particularly when they are unable to bear the loads imposed by their surroundings. Moreover, the presence of these cracks can initiate

* Corresponding author at: School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia.
E-mail address: mazleenda@graduate.utm.my (M. Mazni).

https://doi.org/10.1016/j.aej.2024.02.052
Received 25 October 2023; Received in revised form 23 January 2024; Accepted 22 February 2024
Available online 8 March 2024

the onset of corrosion, which is often irreversible. Consequently, regular and precise damage assessments are crucial. Traditionally, the inspection of concrete structures for the presence of cracks has relied on manual methods. In the context of towering skyscrapers, it is important to note that visual inspections often pose challenges. These challenges include the consumption of significant time and labor, as well as subjectivity in the evaluation process. As outlined in [5], a weakly supervised network has been designed for the segmentation and detection of cracks in asphalt concrete decks. The methodology incorporates autoencoder differentiation of data, accentuating features within unlabeled data, applying k-means clustering for classification, and conducting weakly supervised semantic segmentation on images depicting defects in bridge decks. Through experimentation on a manually annotated dataset featuring six defect types, the proposed approach exhibited better segmentation outcomes than current methods, emphasizing its significance in meeting the essential requirement for accurate damage assessment in concrete structures. The proposed method is highlighted for its importance in accurately assessing damage in concrete structures. Arun Mohan et. al in [6], it presents a comprehensive survey of diverse image processing techniques employed in engineering structures to detect cracks, analyzing each system's approach and highlighting research challenges for further investigation in image processing-based crack detection systems. The study concludes that a significant number of researchers prefer camera-type images, coupled with advanced segmentation algorithms such as threshold and reconstructable feature extraction, for in-depth damage analysis. The study concludes that a significant number of researchers prefer camera-type images, combined with advanced segmentation algorithms like threshold and reconstructable feature extraction, for in-depth damage analysis.

Convolutional Neural Networks (CNNs) stand out as a primary category of neural networks employed in image recognition and classification [7–9]. It represents advanced artificial intelligence systems built on complex multi-layer neural networks. They possess the capability to discern, identify, and categorize objects, alongside the ability to detect and precisely delineate objects within images [7]. A Convolutional Neural Network (CNN) consists of four essential layer types, each serving a distinct purpose within the network's architecture: the convolutional layer, responsible for feature extraction; the pooling layer, which diminishes the spatial dimensions of the feature maps; the activation function layer, applying activation functions; and the fully connected layer. Incorporating the image as input, the convolutional layer combines a limited set of filters, determined by the kernel or filter size, to capture a wide range of relevant features. Following this, the pooling layer reduces the scale of the acquired features, consequently alleviating the computational load on the network. To gain deeper insights from data, researchers have proposed effective models. As presented in [10], CNN is now a vital tool for ensuring the stability of concrete structures through crack diagnosis in civil inspections. A comprehensive review analyzed research papers and categorizing them into methods like classification, segmentation, detection, and hybrids. The article discusses challenges, proposed solutions, and suggests potential research directions in implementing CNN for crack identification.

In the realm of advanced image recognition and classification techniques, researchers have made significant strides in developing models for detecting and classifying structural defects. G. Li et al. introduced "ResNeXt+PP" or ResNeXt with postprocessing [11] for detecting concrete cracks. This model automatically distinguishes between cracks and non-cracks in images, outperforming other methods. Bubryur Kim et al. introduced an exponential Conv2D ResNet exponential model [12] for the accurate classification of wall defects, achieving an F-Score of 99.78%. This model outperformed popular models such as Xception, VGG19, and DenseNet, highlighting its exceptional predictive capabilities in identifying wall defects. Additionally, Zhu et al. [13] demonstrated the use of Inception-V3 networks for preprocessing images with common defects, leveraging transfer learning to enhance classification tasks. In [14], EfficientNetV2 emerged as the top performer, demonstrating outstanding results across all metrics. It excelled in minimizing false positives caused by non-uniform lighting and construction irregularities on surfaces, boasting a remarkable accuracy of 99.6% and precision of 99.3%. Achieving a perfect recall score of 1 showcased its capability to capture all actual crack patches. The model's overall performance was further highlighted by an impressive F1 score of 99.6, illustrating a harmonious balance between precision and recall. Moreover, an inventive wall-climbing Unmanned Aerial System (UAS) in [15] has been designed for crack inspection, marking the first application of a system merging Unmanned Aerial Vehicle (UAV) and wall-climbing robot for Structural Health Monitoring (SHM). This system's unique ability to attach to structures during inspection enables operation in GPS-denied environments and facilitates the acquisition of detailed crack images. Additionally, this characteristic reduces operational difficulties and minimizes the risk of collisions, enhancing the overall efficiency and safety of the inspection process.

Image segmentation methods, particularly region-based approaches such as threshold segmentation, offer valuable tools for detecting contour cracks. Among these methods, OTSU thresholding stands out as the most commonly used technique [16, 17]. OTSU thresholding is a global thresholding method that calculates the image's threshold value based on the intensity histogram, aiming to maximize the variance between foreground and background pixels [18]. This technique proves highly effective when dealing with images featuring uniform backgrounds, where a single threshold value can effectively separate the foreground and background pixels in the resulting binary image [19]. In the context of segmentation using Otsu's method, Convolutional Neural Networks (CNNs) assume a pivotal role in automating the segmentation process. The synergy between CNNs and Otsu's method significantly enhances this process. Integrating Convolutional Neural Networks (CNNs) with Otsu's method not only simplifies the segmentation process but also leads to significant enhancements in both accuracy and efficiency. This synergy is especially advantageous in applications like object classification and image segmentation. Furthermore, this enhances the adaptability of crack recognition, leading to improved accuracy in detection and increased resistance to noise. This approach indicates the potential to address background noise stemming from diverse categories and intensity levels within a unified framework.

This research presents a novel method for image-based crack classification, utilizing transfer learning models rooted in convolutional neural networks (CNNs). Transfer learning is a widely acknowledged technique that leverages pre-trained CNNs models, celebrated for their robustness and high-performance capabilities, in the realm of computer vision. We have explored the use of pre-trained convolutional neural networks to improve classification accuracy for detecting different types of cracks while simultaneously managing concerns related to overfitting and addressing potential class bias, ensuring that our model performs effectively on all categories of cracks. In this study, we utilized pre-trained Convolutional Neural Network (CNN) models, which include MobileNetV2, EfficientNetB7, InceptionV3, and ResNet50. Subsequently, image segmentation was performed using Otsu's method. It is important to clarify that our primary focus was directed towards the classification and measurement of cracks rather than the comprehensive analysis of building materials in exterior walls. Through rigorous training, we achieved optimal results, with the model's performance showing considerable improvement with increased training iterations. Furthermore, we provide an extensive comparative analysis of their performance within the context of crack classification and measurement.

2. Related work

In some instances, prior research has explored the application of transfer learning in the context of crack detection. In [20], Guzmán-Torres et al. took steps to refine a pre-trained CNN using transfer learning, aiming to further improve its accuracy. Their research delved into the evaluation of five pre-trained neural networks, specifically the DenseNet12 model, ResNet50 model, InceptionV3, VGG-16 model, and MobileNet model. These networks each incorporate unique architectural components and assess varying sets of network parameters. As a result of this deep learning (DL) architecture, significant enhancements were observed, most notably with the VGG-16 model, which achieved outstanding results. It achieved a perfect F1-score of 100% and an accuracy of 99.5% when tested on a large dataset. Hajar Zoubir et al., on the other hand, utilized the VGG16 network to optimize a loss function, encompassing both Binary and Multi-Class Cross-Entropy loss [21]. These loss functions quantify the deviation between expected outputs and the actual ground truth through the application of back-propagation. In their study, they trained a VGG16 network using three distinct transfer learning techniques, each affecting different layers of the network. The evaluation of the model's performance in each learning configuration was based on classification metrics, computational time, and its generalization capabilities. The experiments clearly demonstrated that improvements in classification performance were achieved by retraining the classification layers and the final two convolutional layers of the VGG16 network. Optimization methods often incorporate gradient descent optimizers like stochastic gradient descent and adaptive optimizers to adjust the network's learning parameters.

Next, Md. Monirul Islam et al. in [22] utilized four transfer learning models for the experimental setup containing VGG16, ResNet18, DenseNet161, and AlexNet. The lack of training data and overfitting have been addressed using data augmentation techniques as transfer learning, random-resized-crop, random-rotation, color-jitter, and random-horizontal-flip. AlexNet exceeds all other models when it comes to performance indicators, obtaining accuracy rates of 99.90%, precision rates of 99.92%, recall rates of 99.80%, and F1-score of 99.86%. Additionally, it displayed the time taken for each model's training period. In this instance, AlexNet takes first place in less time. In a study conducted by Stamos Katsigiannis et al. [23], transfer learning was harnessed in combination with well-known pre-trained convolutional neural networks to proficiently identify cracks in brickwork masonry. Their experiments yielded remarkable results, achieving 100% accuracy and F1-scores when utilizing models such as MobileNetV2, Inception-ResNetV2, and Xception. Notably, MobileNetV2, due to its compact size, emerged as the most suitable option for applications in handheld mobile devices and wall-climbing robots [24,25]. Furthermore, they conducted a comparative analysis between end-to-end training and training exclusively on the final classification layers. The former consistently outperformed the latter, underscoring the approach's effectiveness in crack detection. In another significant research effort led by Sayyed Bashar Ali and his team [26], a transfer learning model based on MobileNet was developed for the purpose of wall crack detection, delivering an impressive accuracy rate of 99.59%. This research delved into the exploration of various pre-trained CNN models, including MobileNet, InceptionV2, ResNet101, and VGG16, to enhance crack detection, especially in scenarios with limited dataset sizes. These models were adapted by replacing the classification layer with a Soft-Max layer. Evaluation conducted on a relatively small dataset revealed that the lightweight MobileNet model outperformed the other models, with ResNet closely following suit.

3. Material and methods

3.1. Research framework

The research framework devised for classifying and measuring cracks in images consists of five pivotal stages as depicted in Fig. 1. First, "Data Collection" entails the acquisition of diverse crack image datasets. Then, "Data Pre-processing" involves activities like data refinement, resizing, and augmentation. In the "Model Training and Validation" phase, pre-trained CNN models are fine-tuned to enhance their

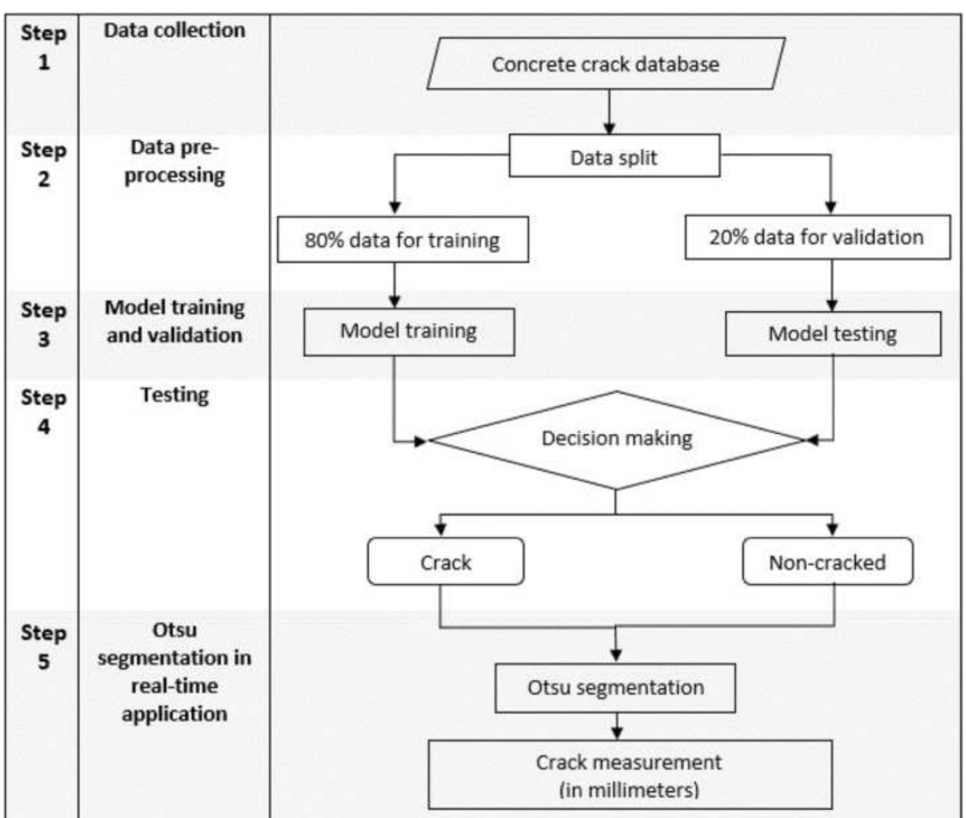

Fig. 1. Crack classification and measurement research framework.

performance in crack detection. This process involves validation, where the dataset is divided into an 80–20 training-validation split. Next, the "Testing" phase rigorously assesses these models by categorizing images into "crack" or "non-cracked". Then, the Otsu method in has been applied in Step 5 to concentrates on implementing real-time crack measurement which measures the size of the detected crack in millimeters. This structured framework guides the research process cohesively, from the initial data acquisition to real-world application, enabling a comprehensive evaluation of pre-trained CNN models and segmentation techniques to boost the effectiveness of crack detection.

3.2. Data collection

The dataset used for this task comprises concrete surface images categorized into two classes: 'negative,' indicating surfaces without cracks, and 'positive,' representing surfaces with cracks. In total, there are 11,435 files, with 9148 designated for the training set and 2287 for validation. This dataset was sourced from the Mendeley Data - Crack Detection website, provided as a valuable resource for crack detection research by Çağlar Fırat Özgenel et al. in [27]. The image dimensions are fixed at 227 ×227 pixels, featuring RGB (Red, Green, Blue) channels. Importantly, no data augmentation techniques, such as random rotation or flipping, were applied. The dataset has been carefully prepared and is now ready for training and evaluating CNNs models for surface crack classification. The images have been uniformly resized to 224 × 224 pixels. This dataset has been divided into two main sets: the training and validation sets, each serving distinct purposes. The training set is employed for model training, while the validation set assists in fine-tuning hyperparameters.

3.3. Selection of object classification framework

There is a wealth of pre-existing convolutional neural networks accessible for the task of classification. An illustrative comparison presenting the manifold architectural alternatives in terms of their accuracy, prediction speed, and model sizes. MobileNetV2 [28,29,30], EfficientNetB7, InceptionV3 [28], and ResNet50 have been selected based on their reputation for achieving a balance between superior accuracy and relatively compact model sizes. These models have gained recognition in the machine learning community for their efficiency in terms of both computational resources and memory footprint. Their suitability for various applications, particularly in scenarios with limited computational power or memory constraints, makes them preferred choices in many cases. Table 1 offers an overview of the principal characteristics inherent to these pre-existing model instances.

Neurons serve as the fundamental components within neural networks (NNs), mirroring the biological neurons in the human brain. They facilitate the transmission of vast amounts of information by applying specific activation functions while considering associated weights. Convolutional neural networks (CNNs) represent a specialized type of NN characterized by multiple hidden layers, making them capable of deep learning. CNNs employ convolution operations between the outputs of various layers, thus earning their name. Activation functions are indispensable regulators of information propagation throughout the neural network and hold a central position in influencing the learning trajectory of the neural network. These operations imbue the network with a non-linear character, facilitating the automated derivation of data-dependent, non-linear attributes from the input images. Some common activation functions include Rectified Linear Unit (ReLU), Sigmoid, Softmax, Hyperbolic tangent activation function (TanH), and Scaled Exponential Linear Unit (SeLU). In this work, we have employed the Sigmoid activation function. Softmax is another popular activation function used in neural networks, particularly in the output layer for classification tasks. Unlike ReLU, which is used in hidden layers to introduce non-linearity, Softmax is primarily used in the output layer to convert raw scores or logits into class probabilities. The Softmax activation function accepts an input consisting of a vector of real numbers and then proceeds to convert this input into a probability distribution that spans across numerous classes. It does this by exponentiating each input value and then normalizing the results. The formula for the Softmax function for a given class (i) is depicted in Eq. (1)

$$Softmax \quad (z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{1}$$

In this context, the symbol Softmax$(z)_i$ denotes the i-th component within the vector representing the output of the softmax function. Similarly, z_i signifies the i-th component of the input vector, where K stands for the overall number of classes involved, and e represents the mathematical constant, Euler's number, which is approximately equal to 2.71828. The Softmax function is employed to ensure that the output values are confined within the range of [0,1], with the additional constraint that they collectively add up to 1. The class with the highest probability is typically chosen as the predicted class. In the context of classification tasks, the Softmax function is frequently employed in the concluding layer of a neural network to generate probabilities for each class, and the predicted category corresponds to the one associated with the highest probability. It's commonly used in conjunction with the categorical cross-entropy loss function for training classifiers. Achieving successful model training and accurate predictions relies heavily on the model's internal parameters. Therefore, it is imperative to carefully choose a suitable optimizer for the purpose of fine-tuning the network's parameters, with the ultimate goal of approximating or achieving the most favorable values. The pivotal element in the reduction of the loss function, adjustment of model parameters, and eventual achievement of convergence is the optimization algorithm. In this specific scenario, our preference is to utilize the Adam algorithm for updating the model's weights. This model configuration specifies that a neural network will be trained with the Adam optimization algorithm, using a learning rate of 0.01. The loss function employed for training purposes will be the Categorical Cross-Entropy, which serves to evaluate the dissimilarity between the predicted class probabilities and the actual class labels. Furthermore, Model Checkpoint callbacks will be utilized to preserve the optimal model weights contingent upon their validation performance throughout the training process. The training regimen will span 30 epochs, during which data will be handled in groups of 32 samples per batch.

Table 1
Assessing parameter count and computational workload in four prominent CNNs architectures: MobileNetV2, EfficientNetB7, InceptionV3, and ResNet50. Notably, MobileNetV2 Excels with the fewest parameters and compact size.

Model	Size (MB)	Parameters (Million)
MobileNetV2	14	3.5
EfficientNetB7	256	66.7
InceptionV3	92	23.9
ResNet50	98	25.6

3.4. Transfer learning

Transfer learning is a well-established method within the field of deep learning, centered on the utilization of knowledge obtained while addressing a particular problem to tackle related issues. This approach proves to be exceptionally advantageous in situations where pre-trained models are harnessed for classification tasks. The process of constructing a Convolutional Neural Network (CNN) from the ground up necessitates extensive training on extensive datasets to guarantee dependable detection results. Consequently, this methodology is frequently avoided, especially when confronted with constraints on the availability of data. In contrast, transfer learning consistently surpasses the alternative, which involves creating a new CNN model with a random initialization and converging it. The fundamental idea behind transfer learning lies in the ability to utilize a model's pre-trained weights from a comprehensive, general-purpose database, retaining these weights up to the final layer, and then appending additional layers as needed. Conversely, building and training a model from the ground up would be computationally expensive, requiring a well-defined architecture and substantial data resources. Transfer learning mitigates these challenges, offering practical solutions for detection and classification problems with relatively small available datasets. Transfer learning models not only train rapidly due to their utilization of pre-trained weights but also tend to deliver superior predictive performance. We were driven to explore pre-trained CNN models, including MobileNetV2, EfficientNetB7, InceptionV3, and ResNet50. All experiments within this paper were conducted on the Windows system using Google Colab, with the following hardware specifications: an Intel(R) Core(TM) i5–3337 U Central Processing Unit (CPU) running at 1.80 GHz, 4 GB of RAM, and Windows 10 Pro as the chosen operating system. Additionally, for the post-training analysis, including the generation of accuracy and loss graphs, as well as confusion matrices, we utilized PyCharm. PyCharm served as a valuable tool for its capabilities in aiding the visualization and comprehensive analysis of our experimental results.

We elected to employ the approach of transfer learning in the training of Convolutional Neural Network (CNN) models, with the primary objective of distinguishing between images labeled as either 'crack' or 'non-crack.' In this particular context, our strategy was founded upon well-established deep learning architectures previously trained on a significantly larger image dataset, demonstrating their efficacy as feature extractors for image classification tasks. Transfer learning models, not only rapidly train by harnessing existing weightings but also exhibit a propensity for delivering enhanced predictive performance. Concretely, we made a deliberate choice to employ four specific CNN architectures that had undergone pre-training using the extensive ImageNet dataset, comprising 1.4 million images categorized into 1000 distinct classes [31].

The chosen architectural models encompass MobileNetV2, EfficientNetV2, InceptionV3, and ResNet50. It is of significance to underscore that, in the pursuit of stabilizing the convolutional base weights across these divergent approaches, we uniformly established the 'trainable' parameter as 'false' for all convolutional base layers within the Keras model implementations. In the context of transfer learning, one practice that we embraced involved the substitution of the fully connected layers, which encompassed the 2D global average pooling layer and a dense layer, complemented by a softmax activation function for the purpose of classification. The inclusion of a softmax activation layer served the vital function of generating probability distributions for the two defined classes, thereby equipping the model with the capability to make precise binary predictions. The global average pooling layer played a pivotal role in dimensionality reduction for the data originating from the convolutional base, while the subsequent dense layer assumed the responsibility of making final predictions based on these derived features. These layers were introduced atop a pre-trained feature extractor, typically a convolutional base, previously honed through extensive training on a substantial dataset. The said pre-trained feature extractor, often constituted by a convolutional neural network (CNN), specialized in the extraction of valuable image features. By replacing the fully connected layers, we tailored the model to our specific classification task, thereby affording us the ability to harness the wealth of knowledge and feature representations encapsulated within the pre-trained model. This adaptability is invaluable, particularly in the context of our specific problem, which pertains to the classification and measurement of structural defects like cracks. A visual depiction of the neural network's architectural design is elucidated in Fig. 2. During the training, the "model checkpoint callback" in TensorFlow has been applied to enable the saving of model checkpoints during training. It offers various filepath that specifies the directory for saving checkpoints, and the usage of "verbose" to controls the alert level when checkpoints are saved. This callback enhances model progress monitoring and management during training.

3.5. Otsu segmentation

A technique for segmenting images based on regions, such as threshold segmentation, can be employed to identify contour cracks. Among these methods, OTSU thresholding, as indicated by Cao et al. in 2021 [32], stands out as a widely adopted approach. This global thresholding method calculates an image-wide threshold value by analyzing the image intensity histogram. It aims to identify the threshold value that maximizes the distinction between foreground and background pixels, as outlined by Chen et al. in 2019 [33]. OTSU thresholding is particularly effective when dealing with images that possess a uniform background, allowing for the clear separation of foreground and background pixels in the resulting binary image, a concept first introduced by Otsu in 1979 [19].

During the assessment of crack dimensions, the program meticulously examines each contour, evaluates the size of individual cracks, and subsequently delineates bounding boxes encompassing these contours. Once the bounding boxes are identified, the code utilizes the Euclidean distance to determine the length and width of each crack [34]. This distance measurement is applied to gauge the length of the cracks. Euclidean distance serves as a means for determining the direct spatial separation between two given points and subsequently transforming it into measurements expressed in millimeters. This is achieved using a specific mathematical formula, as illustrated in Eq. (2).

$$\text{Euclidean distance} = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} \tag{2}$$

Analyzing the contour information within the image frame, it helps determine the width of the cracks using the value of the Euclidean distance. This distance is computed by finding the square root of the sum of squared differences between the x and y coordinates of two points in a

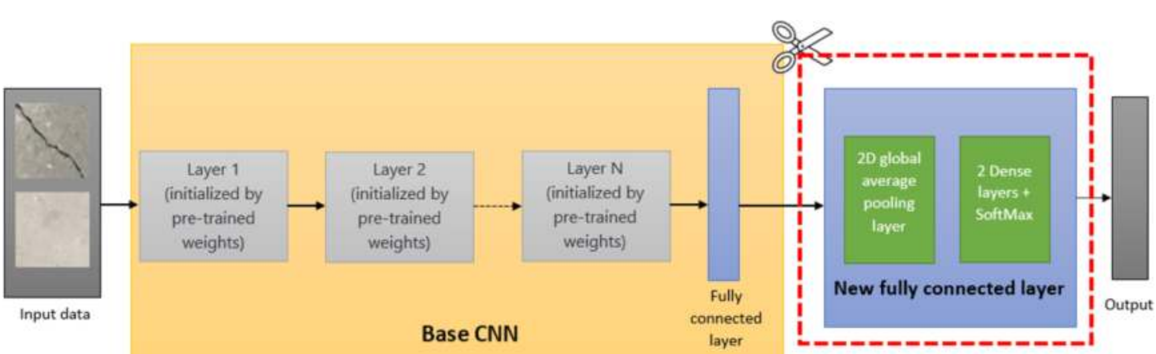

Fig. 2. A summary of the proposed approach. The base CNN's structure (in orange box) varies depending on the chosen pre-trained model's architecture.

two-dimensional space, with these coordinates being the input parameters. Based on the contours detected within a video frame, the technique of pixel per inch has been implemented which use resolution of the video to calculate the maximum width of a crack in millimeters.

3.6. Quality metrics

We have evaluated the effectiveness of our proposed crack detection method by employing standard quality metrics, such as precision, recall, accuracy, and the F1-score, as illustrated in Eqs. (3) through (6); where TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives) correspond to the respective scores derived from the evaluation process. Accuracy, defined as the ratio of true positives (TP) and true negatives (TN) to the sum of TP, TN, false positives (FP), and false negatives (FN). Recall, expressed as TP divided by the sum of TP and FN, measures the method's effectiveness in identifying true positives relative to actual positive cases. Precision, calculated as TP divided by the sum of TP and FP, assesses the method's accuracy in correctly identifying true positives among the predicted positive cases. The F1-score, a harmonic mean of precision and recall, is computed as 2 times the product of precision and recall, divided by the sum of precision and recall. This metric proves valuable in scenarios with imbalanced datasets, offering insight into the algorithm's misclassification rate. These metrics are mathematically defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{F1}_{score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

4. Result and discussion

4.1. Comparative analysis of transfer learning techniques

The graphical representation in Fig. 3 offers a comprehensive view of training and validation accuracy, along with training and validation losses, across various transfer learning (TL) methods, presented in percentages (%). When we examine the top row, a consistent trend emerges in the training accuracy of TL MobileNetV2, TL InceptionV3, and TL ResNet50, with the exception of EfficientNetB7, which achieves the highest accuracy at 88.3%. In comparison, MobileNetV2, InceptionV3, and ResNet50 approach an impressive accuracy level of 99–100%. It's important to note that TL MobileNetV2 stands out with a unique pattern: its training and validation accuracy align closely, showing remarkable consistency and higher accuracy compared to other models. This highlights the robust generalization capabilities of TL MobileNetV2. Conversely, TL EfficientNetB7 displays validation accuracy data with occasional fluctuations across different data points, resulting in an irregular performance pattern during various training epochs. In the second row, both TL MobileNetV2 and TL InceptionV3 models demonstrate rapid decreases in both training and validation loss. However, these losses reach a point of diminishing returns with more training epochs. TL EfficientNetB7 showcases a consistent reduction in loss that closely matches the validation loss, indicating strong agreement between the two. In the case of TL ResNet50 models, a mild phase of overfitting is initially observed, followed by a subsequent reduction in loss values. Notably, some early-epoch fluctuations in validation loss are evident, eventually stabilizing at a lower level.

Table 2 presents the performance metrics, reported in percentage

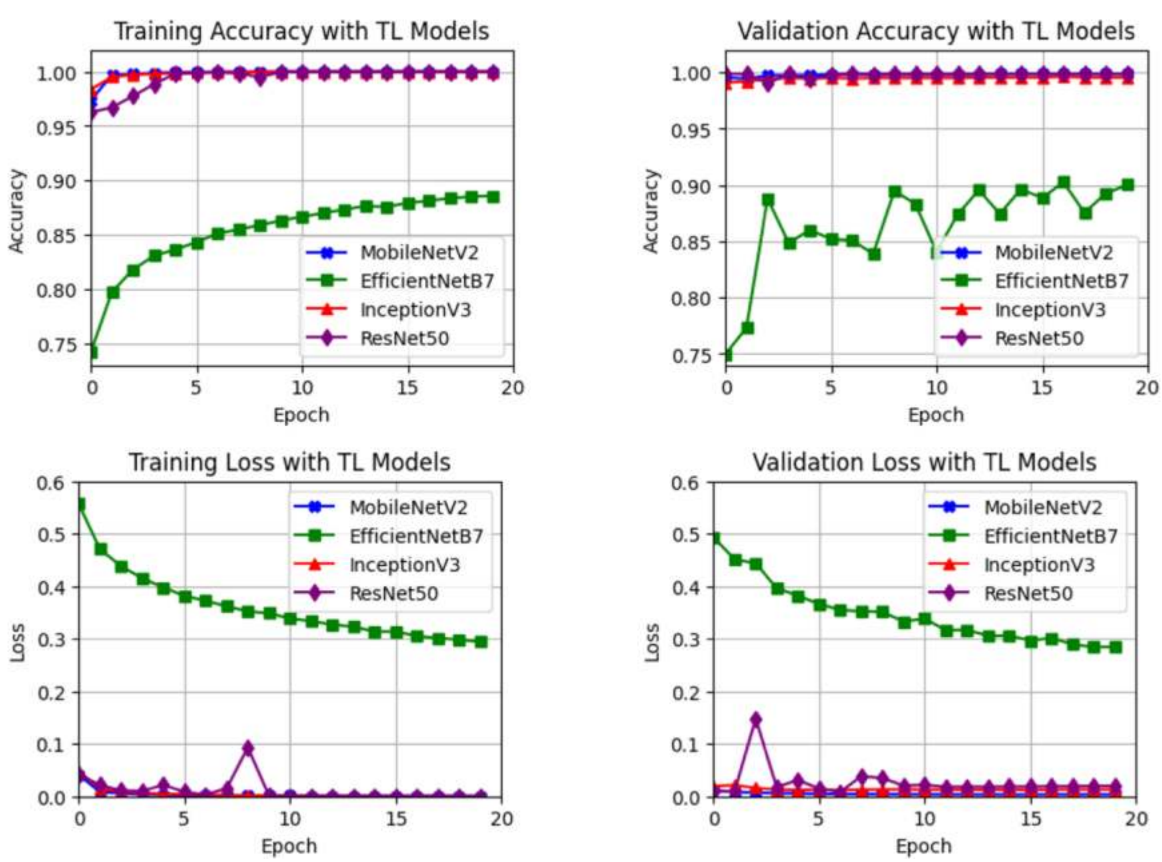

Fig. 3. Training and Validation Accuracy/Loss for Transfer Learning Models (TL) – Percentage (%) Metrics.

Table 2
Evaluating Performance Metrics for Different Pre-trained Transfer Learning CNN Models (%).

Transfer learning Model	Accuracy	Recall	Precision	F1-Score
MobileNetv2	99.87	99.74	100	99.87
EfficientNetB7	90.03	80.57	99.46	89.03
InceptionV3	99.56	99.39	99.74	99.56
ResNet50	99.87	99.91	99.83	99.87

(%), of various pre-trained CNN models with Transfer Learning (TL), including accuracy, precision, recall, and F1-score. Furthermore, Fig. 4 display the confusion matrices for the test set. Based on the results, MobileNetV2 and ResNet50 in particular, emerged as the top-performing models which exhibit comparable accuracy and F1-Score in classifying concrete cracks, with ResNet50 demonstrating a slightly higher recall and MobileNetV2 achieving a perfect precision. However, MobileNetV2 is known for its lightweight architecture, making it efficient in terms of model size and computational resources. ResNet50, on the other hand, is a deeper and more complex model, which may require more computational power and memory. MobileNetV2's lightweight nature may be preferred when resource constraints are a concern, while ResNet50's higher recall could be valuable when ensuring the capture of true positive cases is critical. EfficientNetB7 achieves an accuracy of 90.03%, while InceptionV3 surpasses it with an accuracy of 99.56%. InceptionV3 demonstrates a significant advantage in terms of overall accuracy, indicating its superiority in correctly classifying concrete cracks. EfficientNetB7 is part of the EfficientNet family known for its efficiency in terms of model size and computational resources. In contrast, InceptionV3 is a more complex model, which may require more computational power and memory.

4.2. Crack classification outcomes and performance metrics

In this section, we will explore the results and performance metrics of crack classification using different transfer learning models. Fig. 5 illustrates the crack classification outcomes using the transfer learning models MobileNetV2, EfficientNetB7, InceptionV3, and ResNet50. The displayed percentage represents the "confidence" level of each image and is calculated as a percentage based on the highest SoftMax score from the model's predictions. It indicates the model's confidence in its prediction for a particular image. Based on data MobileNetV2 demonstrates significantly high confidence percentages for every image. For EfficientNetB7, only one image is misclassified (False Negative, FN), where the model should have categorized it as Positive. In contrast, both InceptionV3 and ResNet50 exhibit accurate classification for each image and can be considered efficient.

A performance comparison of our models with the existing similar works has been carried out and presented in Table 3. The most recent deep learning methods for crack-detection are Guzmán-Torres et. al [20], Hajar Zoubir et. al [21], Md. Monirul et. al [22], and Sayyed Bashar Ali et. al [26]. The work in [20] uses a Regularized VGG-16 model with accuracy of 99.38%. It is a CNN model utilizing a binary cross entropy loss function, demonstrated impressive performance, showcasing its efficacy in concrete crack classification. Hajar Zoubir et al. in [21], the VGG16 network, was trained on the proposed dataset following three Transfer Learning (TL) schemes with varying layers, achieved impressive testing accuracy rates of 97.13%, demonstrating superior performance in classifying cracks, efflorescence, and spalling. Md. Monirul Islam et al. [22] addressed training data challenges and overfitting using data augmentation techniques in their experimentation with VGG16, ResNet18, DenseNet161, and AlexNet. AlexNet outshone other models with an impressive accuracy of 99.90%, precision of 99.92%, recall of 99.80%, and an F1-score of 99.86%. Sayyed Bashar Ali et. al [26] developed a transfer learning model based on MobileNet for wall crack detection, attaining an impressive accuracy rate of 99.59%. Their exploration involved adapting various pre-trained CNN models, including MobileNet, InceptionV2, ResNet101, and VGG16, by replacing the classification layer with a SoftMax layer. A performance comparison of our models with the existing similar works has been carried out and presented in Table 3.

4.3. Otsu-based real-time crack measurement

In the past, tools like calipers were employed for the assessment of crack width. Nonetheless, this conventional approach had its constraints, encompassing restricted precision and the substantial time investment required for comprehensive measurements, particularly in scenarios involving a multitude of cracks. In this context, the use of high-quality measurement and image processing techniques becomes critical to obtain precise and useful data for evaluating the structural condition. The safety and strength of the structure depend on this last step in determining the extent of crack damage and the subsequent steps in the maintenance and repair process. In this research, after classifying cracks using transfer learning model, the final step involves segmenting the cracks using the Otsu method. To facilitate this, we leverage modern technology by utilizing an iPhone 14 Pro equipped with the Iriun software. This software, installed on the device, transforms the iPhone 14 Pro into a wireless webcam, allowing direct capture of high-quality images of cracks. Iriun software is a versatile tool that enables users to utilize their smartphones as a wireless webcam for various applications, including capturing images for crack assessment in structural analysis. The software typically consists of a mobile application that can be installed on smartphones and companion software for desktop platforms. The software ensures that the captured images maintain a satisfactory level of quality and resolution, crucial for accurate analysis of structural elements like cracks. High-quality images are essential for precise measurements and assessments in applications such as crack width evaluation. Fig. 6 illustrates images of various concrete surfaces, including (a) rough -textured surfaces with aggregates, (b) scratched surfaces, and (c) smooth surfaces with shadows. The results indicate a "Negative" outcome, signifying "non-cracked" areas. In Fig. 7(a), the results depict cracks along with width measurements in millimeters, revealing a maximum width of 3.8 mm. Subsequently, we performed a comparative test to measure the same crack using a vernier caliper, specifically the Mitutoyo Absolute Digital Caliper which equipped with an Absolute Electromagnetic Induction Linear Encoder, this caliper

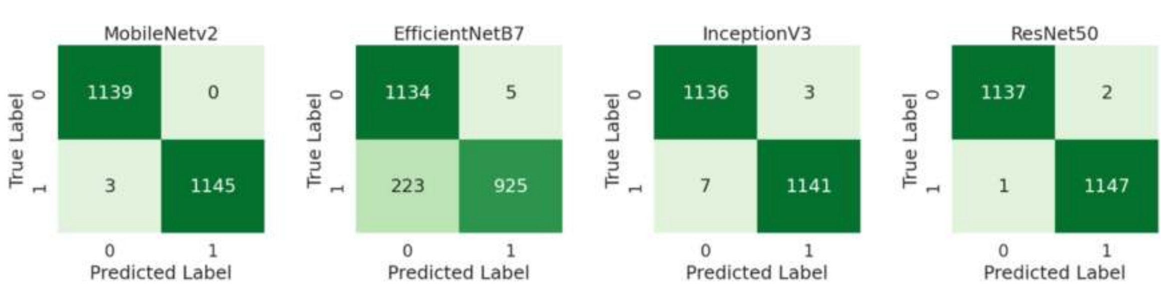

Fig. 4. Confusion matrices for transfer learning CNN models.

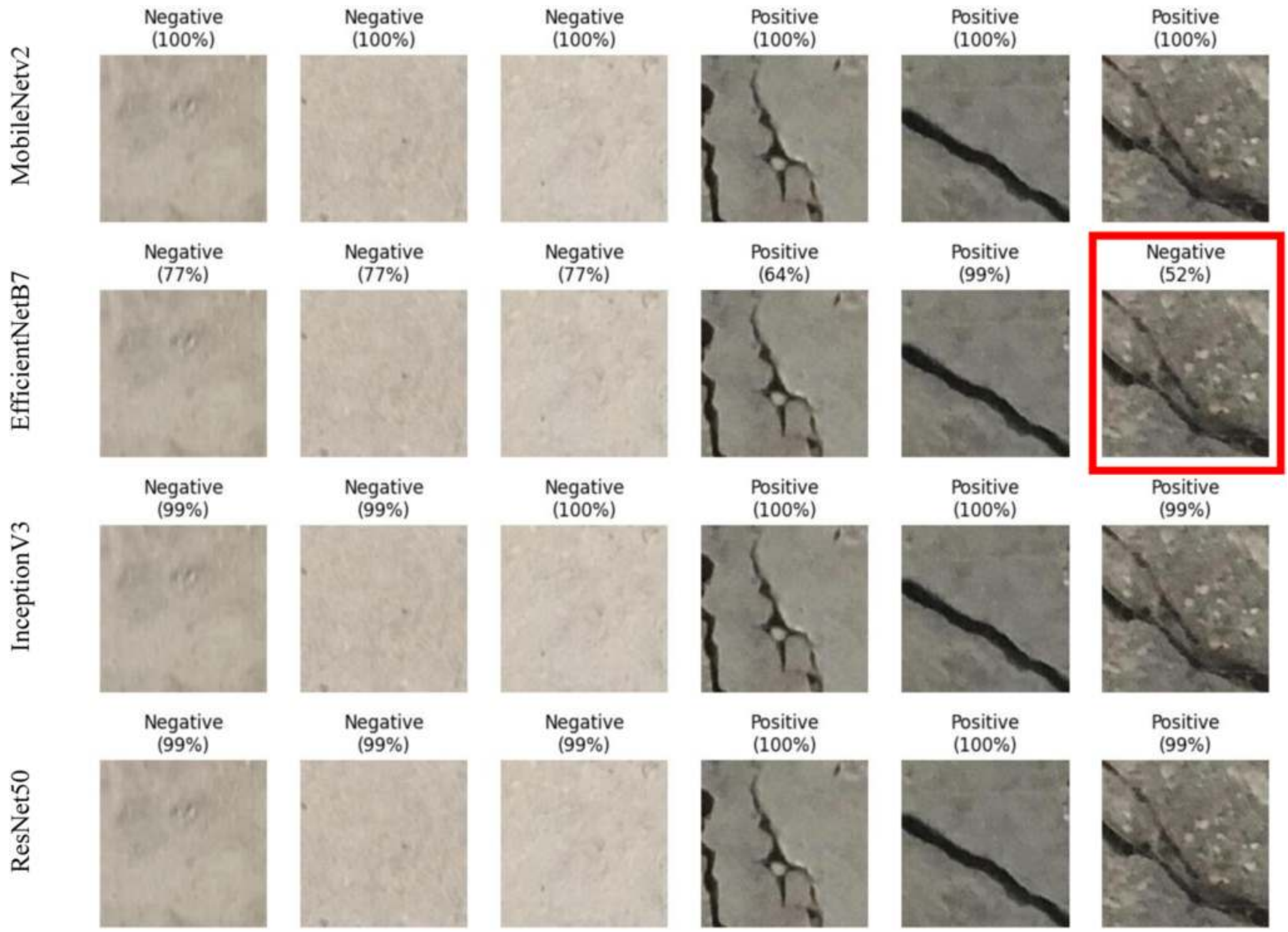

Fig. 5. Sample crack images for with Transfer Learning prediction using CNN models.

Table 3
Performance comparison of the proposed model with the existing works.

Name of author	Model used	Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)
Guzmán-Torres et. al [20]	Regularized VGG-16	99.38	Not specified	Not specified	100
Hajar Zoubir et. al [21]	TL VGG16	97.13	Not specified	Not specified	97.38
Md. Monirul Islam et. al [22]	TL AlexNet	99.90	99.20	99.80	99.86
Sayyed Bashar Ali et. al [26]	TL MobileNet	99.59	Not specified	Not specified	Not specified
Our model*	**TL MobileNetV2**	**99. 87**	**99.74**	**100**	**99.87**

ensures high accuracy with an error margin of ±0.2 mm to ±0.3 mm. Its impressive resolution of 0.01 mm and a Liquid Crystal Display enhance precision and readability. The response speed of this caliper is unlimited, making it an efficient tool for detailed measurements. In Fig. 7(b), the results of the vernier caliper measurements show three readings along the same crack: 3.82 mm, 3.81 mm, and 3.77 mm. Subsequently, the average measurement was calculated to be 3.80 mm, aligning consistently with the width measurement obtained using a smartphone image in real-time. This showcases the reliability and accuracy of the Mitutoyo Absolute Digital Caliper in providing precise crack width measurements for structural evaluation.

5. Limitations

In this pioneering study, we present a novel approach for the classification and measurement of concrete surface cracks, addressing critical challenges in structural assessment. The methodology integrates state-of-the-art transfer learning convolutional neural networks (CNNs) for crack classification and incorporates the Otsu Method for precise crack measurement. We explore the efficacy of four transfer learning models using MobileNetV2, EfficientNetV2, InceptionV3, and ResNet50—in the context of crack classification. Remarkably, MobileNetV2 with transfer learning emerges as a highly effective CNN, exhibiting exceptional accuracy, recall, precision, and F1-score in our experiments. However, despite these promising outcomes, it is imperative to acknowledge certain limitations inherent in our methodology. These limitations encompass factors ranging from the dependency on image quality and challenges in generalization to the sensitivity of crack width calculations and computational resource requirements. By critically examining these limitations, we aim to provide a comprehensive understanding of the potential constraints and areas for further refinement in our proposed classification model and measurement technique.

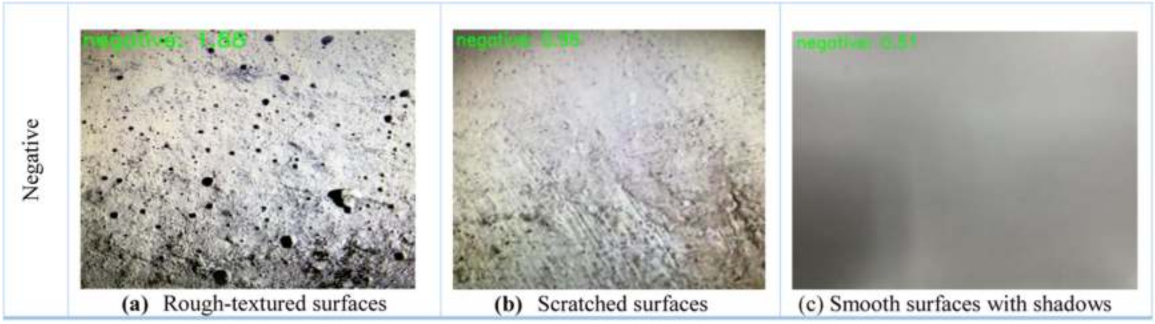

Fig. 6. Concrete surface variations with negative results.

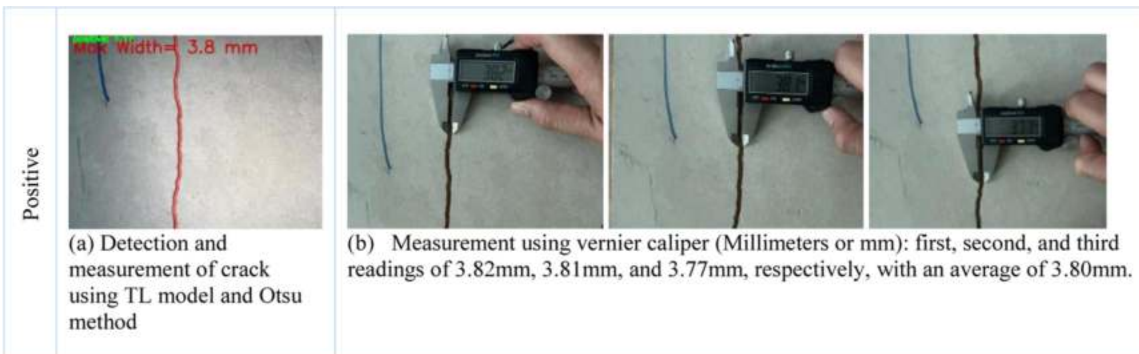

Fig. 7. Concrete surface variations with positive results.

Such insights are crucial for guiding future research endeavors and improving the applicability of our approach in diverse concrete surface scenarios.

1. Dependency on Image Quality
   The accuracy and effectiveness of the classification model and crack measurement heavily rely on the quality of the captured images. Variations in lighting conditions, camera specifications, and image resolution could impact the model's performance.
2. Limited Representation of Real-World Camera Conditions
   Smartphones may vary widely in camera specifications, including sensor types, lens quality, and image processing algorithms. This standardization might not fully encapsulate the diverse imaging scenarios that can occur when employing different smartphone cameras in practical applications. The model's performance in real-world scenarios, where users might employ various smartphones with differing camera capabilities, could be impacted. Factors such as varying lighting conditions, focus issues, and lens distortions inherent in different smartphone cameras may introduce uncertainties not fully accounted for in the current dataset. Future research could benefit from incorporating a more diverse set of images captured under real-world conditions to enhance the model's adaptability to a wider range of smartphone cameras and imaging scenarios.
3. Generalization Challenges
   The presented model's performance is based on specific transfer learning models (MobileNetV2, EfficientNetV2, InceptionV3, and ResNet50) and may face challenges in generalizing to different datasets or diverse environmental conditions, potentially limiting its applicability across various concrete surface scenarios.
4. Sensitivity to Training Data
   The success of transfer learning models is contingent on the representativeness and diversity of the training dataset. If the dataset does not comprehensively cover all possible concrete surface crack variations, the model might struggle to accurately classify and measure novel types of cracks. The dataset used for this task comprises concrete surface images categorized into two classes: 'negative,' indicating surfaces without cracks, and 'positive,' representing surfaces with cracks. In total, there are 11,435 files, with 9148 designated for the training set and 2287 for validation. The success of our transfer learning model is closely tied to the diversity and representativeness of this dataset. The inclusion of various concrete surface conditions, capturing different types of cracks, is crucial for the model to generalize well. A dataset lacking in diversity may hinder the model's ability to accurately classify and measure novel crack types not adequately represented during training. It's worth noting that the careful preparation of the dataset, with fixed image dimensions at 227×227 pixels and RGB channels, provides a standardized basis for model training. However, the absence of data augmentation techniques, such as random rotation or flipping, could potentially limit the model's exposure to a broader range of crack variations. To address the sensitivity to training data, future iterations of this research might consider augmenting the dataset with diverse transformations to enhance the model's ability to handle a wider spectrum of concrete surface crack scenarios during classification.
5. Assumption of Linear Crack Width Calculation
   The method of determining crack width using Euclidean distance assumes a linear relationship between pixel measurements and actual dimensions. However, real-world crack shapes may not always conform to a linear representation, introducing potential inaccuracies in width estimation.
6. Resolution Dependency for Measurement Conversion
   The conversion of pixel measurements to millimeters relies on a pixel per inch technique, which may introduce errors if the video resolution is not consistently maintained. Variations in

resolution could impact the accuracy of the crack width measurements.

7. Limited Validation Scenarios
   The manual experiments using tools like the vernier caliper serve as a validation method; however, they might be limited in representing the full spectrum of concrete surface conditions. The validation should be extended to a broader range of scenarios and environments to enhance the model's robustness.
8. Computational Resource Requirements
   The use of sophisticated transfer learning models may demand substantial computational resources for training and inference. This could pose practical challenges, especially in resource-constrained environments or for researchers with limited access to high-performance computing.
9. Real-Time Implementation Challenges
   While the model demonstrates effectiveness in accuracy, precision, and recall, the real-time implementation on a smartphone may face constraints related to processing speed and memory, potentially limiting its practical utility in dynamic field conditions.
10. Otsu Method Sensitivity
    The Otsu method, while widely used for image thresholding, may be sensitive to variations in lighting and image contrast. This sensitivity could impact the accuracy of crack measurement, especially in situations with inconsistent lighting conditions.
11. Human Subjectivity in Manual Experiments

The manual experiments using tools like the vernier caliper involve a degree of human subjectivity in measurement. Variability in individual operators or interpretation of crack dimensions may introduce uncertainties in the validation process.

These limitations highlight areas where further research and refinement may be necessary to enhance the robustness and applicability of the proposed classification model and measurement technique for concrete surface cracks.

6. Conclusions

In conclusion, this paper introduces a groundbreaking approach to concrete surface crack analysis, featuring a novel classification model and measurement methodology. The core of our model relies on transfer learning Convolutional Neural Networks (CNNs) for classification, seamlessly integrating the Otsu Method for accurate crack measurement. Our experimental setup encompasses four prominent transfer learning models: MobileNetV2, EfficientNetV2, InceptionV3, and ResNet50. Evaluation of their performance is conducted through essential metrics, including accuracy, recall, precision, and F1-score. MobileNetV2, employing transfer learning, emerges as an exceptionally effective CNN, boasting outstanding results with 99.87% accuracy, 99.74% recall, 100% precision, and a 99.87% F1-score in our rigorous experiments. We extend the application of this model to crack measurement, employing the Otsu method. This entails measuring crack sizes on each contour, assessing individual crack dimensions, and delineating bounding boxes around these contours. Utilizing Euclidean distance calculations, we ascertain the width of each crack, enhancing our ability to estimate crack dimensions by scrutinizing contour information within the image frame. The conversion of pixel measurements to millimeters is executed using a pixel per inch technique, accounting for video resolution intricacies.

Beyond the technological advancements highlighted in this research, the potential applications of the developed membranes extend beyond crack assessment. The developed membranes, initially designed for crack assessment, exhibit versatile applications, including pipeline and infrastructure inspection, geotechnical surveys, transportation infrastructure monitoring, and historical structure preservation, demonstrating their potential in various structural evaluation scenarios. In addition to the technological advancements highlighted in this research, the next phase of our investigation will focus on addressing temporal requirements and calculation time, particularly in the context of embedded systems. As we delve into the practical implementation of our groundbreaking approach, it becomes crucial to optimize the computational efficiency of our model for real-time applications.

Declaration of Competing Interest

We, the authors of "An Investigation into Real-time Surface Crack Classification and Measurement for Structural Health Monitoring using Transfer Learning Convolutional Neural Networks and Otsu Method" hereby declare the following regarding potential conflicts of interest in connection with this work:

1. Mazleenda Mazni: No competing interests exist.
2. Abdul Rashid Husain: No competing interests exist.
3. Mohd Ibrahim Shapiai: No competing interests exist.
4. Izni Syahrizal Ibrahim: No competing interests exist.
5. Devi Willieam Anggara: No competing interests exist.
6. Riyadh Zulkifli: No competing interests exist.

We certify that this manuscript represents our original work, has not been published elsewhere, and is not currently under consideration for publication in any other journal. We have all reviewed and approved the final version of this manuscript for submission to Alexandria Engineering Journal.

Acknowledgements

The authors express sincere gratitude for the financial support received for this research. This project was supported by the UTM HIR (Q.J13000.245108G87) and MOHE FRGS (R.J130000.78085F400) research grants. Additionally, the authors acknowledge the support provided by Universiti Teknologi MARA Cawangan Johor (UiTM Johor) through Grant (600-TNCPI 5/3/DDN (01) (001/2021), which played a crucial role in the successful completion of this research.

References

[1] G.L. Golewski, The phenomenon of cracking in cement concretes and reinforced concrete structures: the mechanism of cracks formation, causes of their initiation, types and places of occurrence, and methods of detection—a review, Buildings vol. 13 (3) (2023), https://doi.org/10.3390/buildings13030765.

[2] C.R. Farrar, K. Worden, An introduction to structural health monitoring, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. vol. 365 (1851) (2007) 303–315, https://doi.org/10.1098/rsta.2006.1928.

[3] H.S. Munawar, A.W.A. Hammad, A. Haddad, C.A. Soares, S.T. Waller, Image-based crack detection methods: a review, Infrastructures vol. 6 (8) (2021), https://doi.org/10.3390/infrastructures6080115.

[4] M. Safiuddin, A.B.M.A. Kaish, C.O. Woon, S.N. Raman, Early-age cracking in concrete: causes, consequences, remedialmeasures, and recommendations,", Appl. Sci. vol. 8 (10) (2018) https://doi.org/10.3390/app8101730.

[5] J. Zhu, J. Song, Weakly supervised network based intelligent identification of cracks in asphalt concrete bridge deck, Alex. Eng. J. vol. 59 (3) (2020) 1307–1317, https://doi.org/10.1016/j.aej.2020.02.027.

[6] A. Mohan, S. Poobal, Crack detection using image processing: a critical review and analysis, Alex. Eng. J. vol. 57 (2) (2018) 787–798, https://doi.org/10.1016/j.aej.2017.01.020.

[7] M.M. Taye, Theoretical understanding of convolutional neural network: concepts, architectures, applications, future directions, Computation vol. 11 (3) (2023), https://doi.org/10.3390/computation11030052.

[8] D.O. Brien, J. Andrew Osborne, E. Perez-Duenas, R. Cunningham, Z. Li, Automated crack classification for the CERN underground tunnel infrastructure using deep learning, Tunn. Undergr. Space Technol. vol. 131 (July) (2023) 104668, https://doi.org/10.1016/j.tust.2022.104668, 2021.

[9] Y. Gulzar, Fruit image classification model based on MobileNetV2 with deep transfer learning technique, Sustainability vol. 15 (3) (2023), https://doi.org/10.3390/su15031906.

[10] M. Mazni, A.R. Husain, M.I. Shapiai, I.S. Ibrahim, and D.W. Anggara, "Identification of Concrete Cracks Using Deep Learning Models: A Systematic Review," vol. 8, pp. 1–25, 2024.

[11] G. Li, X. Ren, W. Qiao, B. Ma, Y. Li, Automatic bridge crack identification from concrete surface using ResNeXt with postprocessing, Struct. Control Health Monit. vol. 27 (11) (2020) 1–20, https://doi.org/10.1002/stc.2620.

[12] B. Kim, et al., Deep learning activation layer-based wall quality recognition using Conv2D ResNet exponential transfer learning model, Mathematics vol. 10 (23) (2022), https://doi.org/10.3390/math10234602.

[13] J. Zhu, C. Zhang, H. Qi, Z. Lu, Vision-based defects detection for bridges using transfer learning and convolutional neural networks, Struct. Infrastruct. Eng. vol. 16 (7) (2020) 1037–1049, https://doi.org/10.1080/15732479.2019.1680709.

[14] S.S. Zadeh, S. Aalipour, M. Khorshidi, and F. Kooban, "Concrete Surface Crack Detection with Convolutional-based Deep Learning Models," no. November, 2023, doi: 10.5281/zenodo.10061654.

[15] S. Jiang, J. Zhang, Real-time crack assessment using deep neural networks with wall-climbing unmanned aerial system, Comput. Aided Civ. Infrastruct. Eng. vol. 35 (6) (2020) 549–564, https://doi.org/10.1111/mice.12519.

[16] J. Cao, Research on crack detection of bridge deck based on computer vision, IOP Conf. Ser. Earth Environ. Sci. vol. 768 (1) (2021), https://doi.org/10.1088/1755-1315/768/1/012161.

[17] Q. Cao, L. Qingge, P. Yang, Performance analysis of otsu-based thresholding algorithms: a comparative study, J. Sens. vol. 2021 (2021) 4896853, https://doi.org/10.1155/2021/4896853.

[18] B. Chen, X. Zhang, R. Wang, Z. Li, W. Deng, Detect concrete cracks based on Otsu algorithm with differential image, J. Eng. vol. 2019 (23) (2019) 9088–9091, https://doi.org/10.1049/joe.2018.9191 (Dec).

[19] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. vol. 9 (1) (Jan. 1979) 62–66, https://doi.org/10.1109/TSMC.1979.4310076.

[20] J.A. Guzmán-Torres, M.Z. Naser, F.J. Domínguez-Mota, Effective medium crack classification on laboratory concrete specimens via competitive machine learning, Structures vol. 37 (January) (2022) 858–870, https://doi.org/10.1016/j.istruc.2022.01.061.

[21] H. Zoubir, M. Rguig, M. El Aroussi, A. Chehri, R. Saadane, G. Jeon, Concrete bridge defects identification and localization based on classification deep convolutional neural networks and transfer learning, Remote Sens. vol. 14 (19) (2022), https://doi.org/10.3390/rs14194882.

[22] M.M. Islam, M.B. Hossain, M.N. Akhtar, M.A. Moni, K.F. Hasan, CNN based on transfer learning models using data augmentation and transformation for detection of concrete crack, Algorithms vol. 15 (8) (2022), https://doi.org/10.3390/a15080287.

[23] S. Katsigiannis, S. Seyedzadeh, A. Agapiou, N. Ramzan, Deep learning for crack detection on masonry façades using limited data and transfer learning, J. Build. Eng. vol. 76 (March) (2023) 107105, https://doi.org/10.1016/j.jobe.2023.107105.

[24] R. Zulkifli, A.R. Husain, I.S. Ibrahim, M. Mazni, N.H.A.M. Fauzan, in: Analysis of the Hybrid Adhesion Mechanism of the Wall Climbing Robot BT - Control, Instrumentation and Mechatronics: Theory and Practice, Springer Nature Singapore, Singapore, 2022, pp. 155–169, https://doi.org/10.1007/978-981-19-3923-5_14.

[25] M. Mazni, A.R. Husain, M.I. Shapiai, I.S. Ibrahim, R. Zulkifli, D.W. Anggara, Real-Time Crack Classification with Wall-Climbing Robot Using MobileNetV2, in: F. Hassan, N. Sunar, M.A. Mohd Basri, M.S.A. Mahmud, M.H.I. Ishak, M. S. Mohamed Ali (Eds.), Methods and Applications for Modeling and Simulation of Complex Systems, Springer Nature Singapore, Singapore, 2024, pp. 319–328, https://doi.org/10.1007/978-981-99-7240-1_25.

[26] S.B. Ali, R. Wate, S. Kujur, A. Singh, and S. Kumar, "Wall Crack Detection Using Transfer Learning-based CNN Models," 2020 IEEE 17th India Council International Conference, INDICON 2020, 2020, doi: 10.1109/INDICON49873.2020.9342392.

[27] F. Özgenel and A.G.önenç Sorguç, "Performance comparison of pretrained convolutional neural networks on crack detection in buildings," ISARC 2018 - 35th International Symposium on Automation and Robotics in Construction and International AEC/FM Hackathon: The Future of Building Things, no. Isarc, 2018, doi: 10.22260/isarc2018/0094.

[28] W. Qayyum, A. Ahmad, N. Chairman, and A. Aljuhni, "Evaluation of GoogLenet, Mobilenetv2, and Inceptionv3, pre-trained convolutional neural networks for detection and classification of concrete crack images," 1st International Conference on Advances in Civil & Environmental Engineering, University of Engineering & Technology Taxila, Pakistan, no. March, pp. 2–3, 2022, [Online]. Available: ⟨https://www.researchgate.net/publication/359615441⟩.

[29] Y.C. Hum, et al., The development of skin lesion detection application in smart handheld devices using deep neural networks, Multimed. Tools Appl. vol. 81 (29) (2022) 41579–41610, https://doi.org/10.1007/s11042-021-11013-9.

[30] N.Van Hieu, N.L.H. Hien, Automatic plant image identification of Vietnamese species using deep learning models, Int. J. Eng. Trends Technol. vol. 68 (4) (2020) 25–31, https://doi.org/10.14445/22315381/IJETT-V68I4P205S.

[31] O. Russakovsky, et al., ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. vol. 115 (3) (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y.

[32] J. Cao, Research on crack detection of bridge deck based on computer vision, IOP Conf. Ser. Earth Environ. Sci. vol. 768 (1) (May 2021) 012161, https://doi.org/10.1088/1755-1315/768/1/012161.

[33] X. Chen, J. Li, S. Huang, H. Cui, P. Liu, Q. Sun, An automatic concrete crack-detection method fusing point clouds and images based on improved otsu's algorithm, Sensors vol. 21 (5) (2021) 1–19, https://doi.org/10.3390/s21051581.

[34] M. Flah, A.R. Suleiman, M.L. Nehdi, Classification and quantification of cracks in concrete structures using deep learning image-based techniques, Cem. Concr. Compos. vol. 114 (May) (2020) 103781, https://doi.org/10.1016/j.cemconcomp.2020.103781.