APPLICATIONS OF
MODELLING AND SIMULATION
http://arqiipubl.com/ams eISSN 2600-8084 VOL 9, 2025, 154-163

A Novel Approach to Multi-Class Crack Classification: ResNet50 Enhanced with Squeeze-and-Excitation Blocks

Mazleenda Mazni[1,2], Abdul Rashid Husain[1*], Devi Willieam Anggara[3], Mohd Ibrahim Shapiai[4], Riyadh Zulkifli[1] and Munirah Onn[5]

[1]Faculty of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Malaysia
[2]Department of Dynamics, Control & System Engineering, Faculty of Mechanical Engineering, Universiti Teknologi MARA Johor, Malaysia
[3]School of Electrical and Informatics Engineering, Institute of Technology Bandung, Bandung, Indonesia
[4]Centre for Artificial Intelligence and Robotics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
[5]Faculty of Applied Sciences, Universiti Teknologi MARA Johor, Malaysia

[*]Corresponding author: abrashid@utm.my

Submitted 07 November 2024; Revised 27 March 2025; Accepted 15 April 2025; Available online 08 May 2025.

Abstract: This study focuses on introducing a novel multi-class crack classification framework for concrete structures by integrating Squeeze-and-Excitation (SE) blocks into the ResNet50 architecture. Unlike conventional models, which lack feature recalibration capabilities, the embedded SE block adaptively enhances channel-wise feature importance, significantly improving crack detection performance under complex conditions such as lighting variations and background noise. The proposed model is trained on a comprehensive dataset of 7,147 images, derived from the Özgenel dataset, the Concrete Crack Conglomerate Dataset, and additional images collected from buildings at Universiti Teknologi Malaysia (UTM). Experimental results demonstrate that ResNet50-SE achieves a remarkable accuracy of 99.88%, outperforming the standard ResNet50 (93.25%) and surpassing existing multi-class crack classification models. Furthermore, the optimized architecture ensures computational efficiency, making it well-suited for real-time structural health monitoring applications. These findings validate the effectiveness of SE-enhanced architectures in improving feature extraction and classification accuracy, marking a significant advancement in automated crack detection.

Keywords: Classification accuracy; Concrete structures; Crack classification; Feature extraction; Squeeze-and-Excitation.

1. INTRODUCTION

Concrete structures such as buildings [1]-[6], bridges [7]-[10], and roads [11]-[13] are prone to cracking over time due to environmental exposure, thermal changes, overloading, and material fatigue. These cracks, if left unaddressed, can threaten structural integrity, reduce service life, and significantly increase maintenance costs. Early detection is vital in mitigating risks and prolonging the useful life of critical infrastructures. However, traditional practices of manual inspection are invariably laborious, liable to human error, and impractical for assessing large or complex structures within a limited timeframe [14, 15] . The development of automated systems using computer vision and deep learning have provided innovative solutions for crack detection. Models such as convolutional neural networks (CNNs) offer high accuracy in recognizing patterns within crack images, significantly outperforming methods like thresholding and filtering [16]. Noisy images and variable lighting conditions on backgrounds create problems for crack detection and result in loss of overall accuracy [17].

While the work done has shown improvements, challenges still persist in terms of dataset size, image quality, and effective pre-processing methods to further enhance model robustness and better generalization. This work develops an improved variant of ResNet50 networks with embedded Squeeze-and-Excitation (SE) blocks to address these challenges. The goal is feature learning enhancement and boosting in classification accuracy while keeping computational efficiency intact such that the solution is workable at large scales for actual civil engineering applications. This research corresponds to the advancement of automated crack detection technologies for the development of reliable and scalable structural health monitoring systems, which are of essence in assuring timely maintenance and support the longevity of infrastructure.

2. RELATED WORK

Numerous studies in literature have investigated binary crack classification, concentrating on differentiating between cracked and intact surfaces. Bandi *et al.* [18] utilized a dataset containing 56,000 images from Utah State University and applied the ResNet-50 model, achieving high accuracy despite obstacles such as shadows. Mazni *et al.* [19] used Transfer Learning (TL) with models such as MobileNetV2, EfficientNetV2, and ResNet-50, with MobileNetV2 achieving 99.87% accuracy and precise millimeter-level crack measurement using the Otsu method and Euclidean distance. Next, Islam *et al.* [20] also explored crack classification by type, leveraging TL models like AlexNet, which achieved top results on the CCIC dataset. Further validation using the different dataset from Kaggle confirmed the effectiveness of these transfer learning models for detecting and classifying cracks across multiple datasets.

Beyond binary classification, multi-class crack classification has gained attention, with researchers categorizing cracks by severity or type. Wibowo *et al.* [21] focused on severity-based classification by using TL with VGG16 and ResNet-50, combined with ANN and kNN classifiers, to classify wall cracks into light, medium, and severe levels. ResNet-50 paired with kNN provided the most consistent performance. Anggara *et al.* [22] employed a wall-climbing robot with imaging technology to automate concrete crack detection and classification. The models utilized for crack orientation classification are based on Support Vector Machines (SVM), Random Forest, K-Nearest Neighbors, and Decision Tree, with the shape threshold handled with OTSU. Not crack, Crocodile, Transverse, and Longitudinal are assigned as different orientations of cracks. Comparatively, SVM achieved the best accuracy in one-class classification at 99.50% and in multi-class classification at 87%, against other models such as Random Forest, Decision Tree, and K-Nearest Neighbors.

Recently, robotic-based crack detection systems have gained traction for real-time structural health monitoring. Recently, robotic-based crack detection systems have gained traction for real-time structural health monitoring. The study by Mazni *et al.* [23] introduced an autonomous wall-climbing robot integrated with MobileNetV2 to enhance crack classification performance under varying environmental conditions. Their findings highlight the advantages of robotic-assisted crack inspection, particularly in reducing manual labor, increasing inspection efficiency, and covering complex infrastructure surfaces. Additionally, Zulkifli *et al.* [24] proposed a hybrid adhesion mechanism to improve climbing robots' stability and adaptability, ensuring reliable data collection during real-time inspection tasks. While robotic-based approaches improve mobility and automation, deep learning models like ResNet50 and ResNet50-SE provide enhanced feature extraction and classification accuracy. Future research could explore hybrid solutions, where robotic platforms collect image data while deep learning models like ResNet50-SE perform high-accuracy classification, offering a comprehensive approach to automated crack detection. Integrated Systems in Educational Applications Integrated systems are also being used to enhance the practical understanding of control processes within the field of industrial automation. As an example, [25] created a training platform based on Programmable Logic Controllers (PLCs) integrated with visualization environments as well as real input and output modules which provide opportunities for participants to learn practical experience and immediate feedback on industrial automation knowledge. This pedagogical strategy supports experiential learning while already hinting towards infrastructure applications of structural health monitoring, where real-time monitoring and control requires integration of hardware with the software, while timely and reliable crack detection is crucial for maintenance and safety.

The reviewed works on both binary and multi-class crack classification have shown exemplary advancement in models and techniques. Binary classification was focused on classifying surfaces between cracked and non-cracked, proposing techniques like Transfer Learning, deep learning models ResNet-50, MobileNetV2, and AlexNet with high accuracy on various datasets. Research in multi-class classification predominantly focuses on the identification of either crack severity or crack orientation, proposing models such as VGG16, ResNet-50, and various machine learning classifiers. This study introduces an improved classification framework that utilizes ResNet50 combined with SE blocks to tackle multi-class crack classification issues. This methodology seeks to enhance model adaptability for various crack types and classification scenarios, rectifying a deficiency noted in previous research that predominantly emphasized binary classification accuracy or severity-based categorization. The proposed framework enhances feature extraction by integrating SE blocks with ResNet50, optimizing it for precise crack classification, essential for effective structural health monitoring applications.

3. PROPOSED CLASSIFICATION FRAMEWORK

The proposed model integrates ResNet50 with a SE block, as illustrated in Figure 1. The ResNet50 architecture comprises numerous residual units, each consisting of three convolutional layers, batch normalization, and activation functions. These units facilitate the model's acquisition of intricate features by preserving information throughout layers and mitigating the vanishing gradient problem in deep networks. The ResNet50 backbone, comprising 53 convolutional layers, offers a robust foundation for feature extraction.

The SE block is incorporated to recalibrate the model's emphasis on significant features, functioning through three stages: Squeeze, Excitation, and Rescale. A Global Average Pooling (GAP) layer first compresses spatial information into a compact representation while retaining channel details. Subsequently, this compressed data undergoes two dense layers for "excitation," during which each channel is re-weighted according to its significance. Finally, these modified features are 'rescaled' across input channels, allowing the model to focus on relevant features crucial for crack classification.

Figure 1 illustrates the SE block, which is highlighted in orange, featuring an additional dropout layer following to the initial ReLU-activated dense layer. The dropout layer, which is not typically included in the SE block, enhances model generalization by randomly deactivating some input units during training, encouraging the model to learn general patterns rather than memorizing specific training data. This method mitigates overfitting, particularly beneficial for varied crack images exhibiting diverse textures, lighting, and shapes. As a result, the model becomes more robust and effective in handling new data, which is crucial for practical applications.

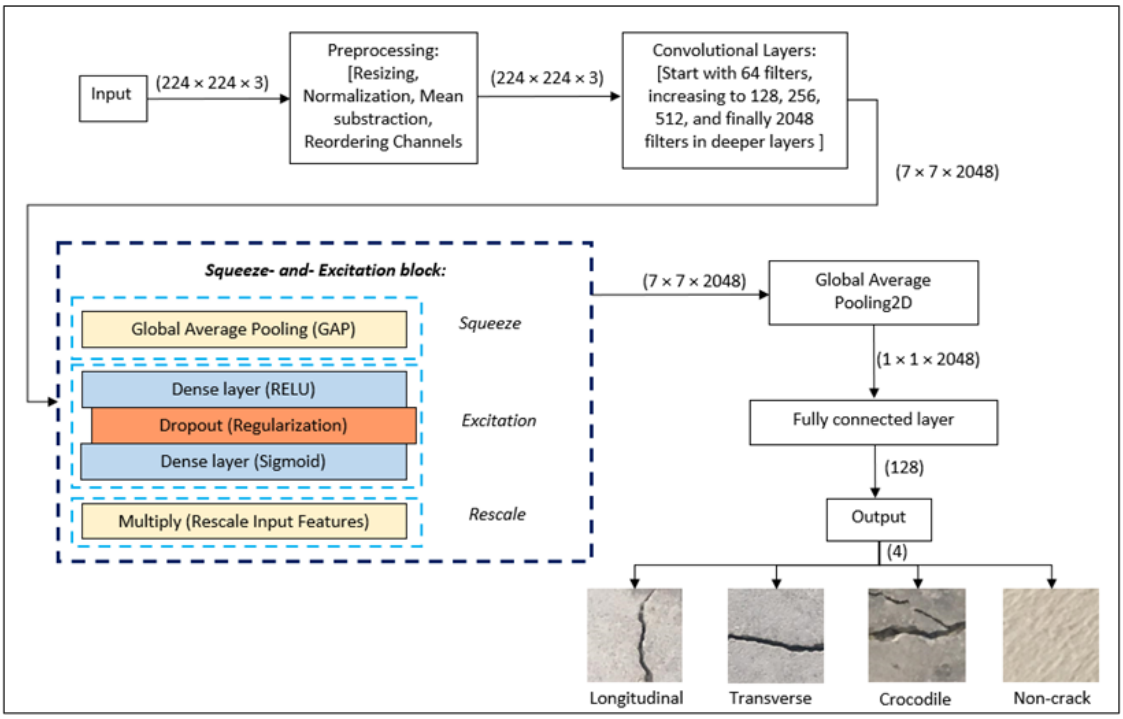

Figure 1. Block diagram of the proposed ResNeT50-SE model.

The comprehensive model configuration comprises 53 convolutional layers derived from ResNet50, three fully connected layers, a global average pooling layer, a dropout layer, and a SoftMax layer for multi-class classification. The SoftMax layer assigns probabilities to four crack classes: Longitudinal, Transverse, Crocodile, and Non-crack, ensuring accurate and clear predictions. This design meticulously balances depth and efficiency, facilitating comprehensive feature extraction without overburdening computational resources. Utilizing SE blocks enables the model to concentrate on critical features, thereby improving accuracy in intricate crack classifications. This renders the model appropriate for implementation in structural health monitoring systems where accuracy and dependability are essential.

3.1 Paragraphs Squeeze and Excitation Blocks

The authors of this research have improved ResNet50's feature extraction capabilities by integrating SE blocks into it. SE blocks adaptively re-weight the feature maps according to their relative relevance, improving channel-wise feature recalibration. The SE block acts in three major steps: Squeeze, Excitation, and Rescale. For this work, SE blocks have been used over the ResNet50 backbone to enhance the representational power of the latter one while focusing on the most important features that are required for crack detection.

3.1.1 Squeeze Operation

The input feature map $\cup \in R^{H \times W \times C}$ undergoes GAP to generate a channel descriptor that captures global spatial information. This results in a tensor with a size of $1 \times 1 \times C$, where C represents the number of input channels. This operation is represented mathematically in Equation (1) as:

$$z_C = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j), \text{for } c = 1, \ldots, C \tag{1}$$

where H and W are the height and width of the input feature map, $u_c(i,j)$ represents the feature value at position (i,j) in channel c and z_C is the squeezed output, resulting in a $1 \times 1 \times C$ tensor.

3.1.2 Excitation Operation

The output of the squeeze step $z \in R^{H \times W \times C}$, is passed through two fully connected Dense (fully connected) with ReLU and sigmoid activations, respectively. The intermediate layer reduces the channel dimensionality by a factor (ratio) to learn compact channel dependencies. A dropout layer is also included for regularization to prevent overfitting. The mathematical formulation can be expressed in Equation (2) as:

$$s = \sigma\big(W_2 \delta(W_1 z)\big) \tag{2}$$

where $W_1 \epsilon R^{\frac{C}{r} \times C}$ and $W_2 \epsilon R^{C \times \frac{C}{r}}$ are the weight matrices for the dense layers, δ represents the ReLU activation function, and σ represents the sigmoid activation function. r is the reduction ratio, which controls the capacity and complexity of the SE block and $s \in R^{1 \times 1 \times C}$ denotes the output after the excitation operation, which provides adaptive channel weights.

3.1.3 Rescale Operation

The recalibrated weights are then applied to rescale the original input feature maps U as Equation (3):

$$\widehat{U}_c = s_c \cdot U_c, \text{for } c = 1, \cdots, C \tag{3}$$

where s_c is the scalar weight applied to each channel c and $\widehat{U}_c$ represent the rescaled output feature map.

3.2 Model Architecture

The proposed classification model is based on the pre-trained ResNet50 architecture, which has demonstrated robust feature extraction capabilities for a wide range of tasks. The SE block is combined at the end of ResNet50 layers to enhance the ability for crack type discrimination. The architecture is discussed as represented in Table 1.

The ResNet50 network, pre-trained on the ImageNet dataset, is utilized as the feature extractor in this study. The model's input shape is defined as (224, 224, 3), corresponding to the size of the crack images employed. The ResNet50 model is used without its final classification layer and is initially frozen to maintain the pre-trained weights during the early stages of training. The input images undergo pre-processing tailored to ImageNet standards, where pixel values are normalized to meet the input requirements of ResNet50. After passing through ResNet50, the images proceed to the Squeeze-and-Excitation (SE) block. This block begins with a Global Average Pooling (GlobalAveragePooling2D) layer, followed by a Reshape operation that compresses spatial dimensions to 1x1. This "squeeze" operation condenses spatial information across channels to generate a global descriptor. Then, "excitation" is performed by means of two dense layers: First, the dense layer consists of 128 units, derived by dividing the number of input channels (2048) by the reduction ratio (r = 16). This configuration creates a bottleneck to learn compact channel dependencies and uses ReLU activation.

The second dense layer consists of fewer filter units, which restore the dimensionality, and it uses a sigmoid activation function to produce the weights for recalibration of each channel, allowing the SE block to highlight the most informative channels. That is the generic approach of the SE block, for learning channel-wise scaling factors to amplify the most critical features. Besides, a dropout layer is added after the first dense layer to regularize the excitation process of the model so that better generalization and mitigation of overfitting can occur. Finally, it applies the gained scaling factors via Multiply operation to the original input tensor, which actually recalibrates the feature maps to focus on the most informative channels. This recalibration provided from the SE block enhances the classification performance of ResNet50 by emphasizing the important features extracted therein.

Table 1. The detailed construction and settings of the ResNet50-SE blocks.

Layer	Kernel Size	Filter	Activation	Output Shape
Input	-	-	-	224 × 224 × 3
ResNet50 Backbone	-	-	-	7 × 7 × 2048
Squeeze-and-Excitation	1 × 1	-	ReLU, Sigmoid	7 × 7 × 2048
Global Average Pooling	-	-	-	1 × 1 × 2048
Dense Layer	-	128	ReLU	128
Output Layer	-	4	Softmax	4

3.3 Training Strategy

Figure 2 presents the difference in training strategy between the ResNet50-SE model and the conventional ResNet50 model, which is without fine-turning. This makes the phase-by-phase training of the ResNet50-SE model very different from that of its vanilla version, with added features like dynamic adjustment of learning rates, early stop, and dropout regularisation in the SE block, hence more advanced and efficient. Advanced techniques are used to provide more refinement and adaptiveness in the training strategy adopted by ResNet50-SE as compared to ordinary techniques. This is achieved with the two-phase training approach where, in the first phase, all the layers in the pre-trained ResNet50 base are frozen. The model thereby gets to learn general features from the ImageNet dataset. In the second phase, fine-tuning is implemented whereby the last 20 layers of ResNet50 are unfrozen. This step-by-step process retains crucial pre-trained knowledge while allowing the model to adapt to the task at hand; hence, furthermore specialized learning becomes possible if fine-tuning is carried out.

Being able to make dynamic adjustments to the learning rate is a big part of model refinement. The *ReduceLROnPlateau* callback will automatically decrease the learning rate in correspondence with whenever the validation loss has stopped decreasing, helping to converge easier and more effectively with less overshooting of optimal solutions and to encourage more precise learning as the model reaches its final stages of training. Early stopping is also designed to avoid overfitting, and thus generalization capability of the model is enhanced further. If the accuracy of validation does not improve in 10 continuous epochs, training shall stop. This approach assists not only in saving computational resources but also keeps the model away from overfitting toward the training data and hence maintains the ability to generalize over unseen data well.

Finally, dropout regularization is applied in the SE block of the model to prevent overfitting by randomly deactivating connections during training. This forces the model to learn more generalized features, improving its robustness and ability to adapt to unseen data. Due to enhanced regularization, the generalization capability of the model increases, and better performance is exhibited on the validation set. It includes phased training, dynamic adjustment of the learning rate, early stopping together with dropout regularization within the SE block, which turns ResNet50-SE more exquisite and robust in its strategy for training compared to a vanilla ResNet50. A strategy will, therefore, be highly efficient and ensure better performance in terms of generalization and convergence.

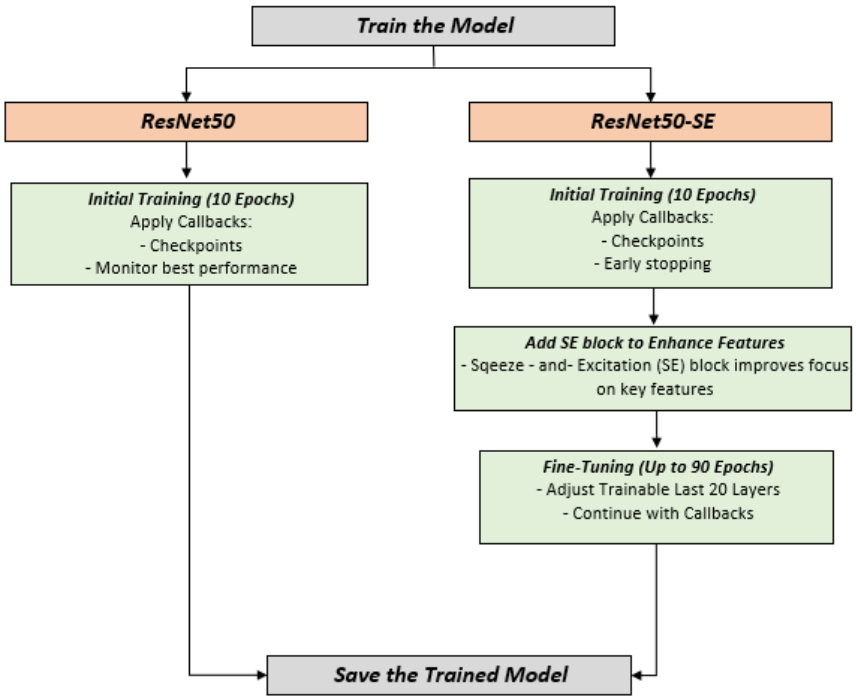

Figure 2. Comparison of training strategies: ResNet50 vs. ResNet50-SE.

4. EXPERIMENTAL SETUP, RESULTS AND ANALYSIS

4.1 Training Strategy

It also included a proper mix of datasets in order to show the robustness for accurate classification of concrete crack images in training, validation, and testing. These phases used a cumulative total of 7,147 images, which were obtained by combining 5,074 images from three sources-the dataset of building cracks around Universiti Teknologi Malaysia (UTM), the dataset from Özgenel *et al.* [26], and the Concrete Crack Conglomerate Dataset [27] that encompasses its sub-datasets, viz. DeepCrack [28], Eugen Miller, and Rissbilder. This combination provided the necessary diversity to train the model effectively, enabling it to recognize various types of concrete cracks. Results at the validation phase, which did not impact the models' parameters during the development, were made on 1,273 images: 800 images from the rest of the Özgenel dataset, and the remaining ones from the Concrete Crack Conglomerate Dataset, including sub-datasets DeepCrack, Eugen Miller, Rissbilder, and the UTM building dataset. The diversity of this dataset will make possible a thorough test of the model capability to properly classify images with concrete cracks of multiple sources. Out of these, 800 images were used for the testing phase. These test images were a mix obtained from the validation set, each to be used in testing the performance of the final model that had been trained. What happens with this walnut mix from the validation set is that it ensures the model's real-life testing of generalizing ability on new, unseen data.

4.2 Performance Analysis

To evaluate the performance of our classifier, several key metrics are needed to understand the effectiveness in recognizing compound characters. These metrics serve as clues about correctness, including samples that are correctly identified and classified by the model. The confusion matrix is a crucial tool in this process, as it provides an overview of the predictions resulting from the classifier measured against the actual labels, thereby assisting the visualization of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Accuracy represents the proportion of correctly classified samples out of the total dataset. Precision measures the proportion of correctly predicted positive instances among all instances classified as positive. Recall indicates the proportion of actual positive instances correctly identified by the classifier.

Being the harmonic mean of Precision and Recall, the average of Accuracy, Precision, Recall, and F1-score for the whole classifier was calculated. This is reflected in Equations (4) to (7). True Positive (TP): The classifier correctly identifies a compound character as correct. True Negative (TN): The classifier accurately predicts an incorrect compound character as incorrect. False Positive (FP): The classifier incorrectly predicts a compound character as correct. False Negative (FN): The classifier fails to identify a correct compound character, predicting it as incorrect. These metrics collectively provide a comprehensive assessment of the classifier's performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$\mathrm{F1}_{score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

All simulations were conducted on a system featuring an AMD Ryzen 5 7535HS processor with a 3.30 GHz clock speed. The setup includes 24 GB of RAM, runs on a 64-bit Windows 11 operating system, and is equipped with an NVIDIA GeForce RTX graphics card, which enhances graphical and computational performance for a range of tasks. The metric analysis shows that the ResNet50-SE model outperforms the ResNet50 model in all major indicators (Table 2). The precision increased significantly, such as surging from 87.95% to 99.50% for the Longitudinal class in ResNet50-SE, and reaching a perfect score of precision 100% for Transverse, Crocodile, and Non-crack classes, verifying the correctness of the positive identification of different crack types by the current model. Recall also evidences significant gains, increasing from 98.50% to 100.00% for Longitudinal and from 89.00% to 100.00% for Transverse, showing further the improved power of the ResNet50-SE model in detecting all positive instances. For the class of Crocodile, precision and recall even leap to 100% from 90.05% and 86.00%, respectively, an improvement regarding proper classification. These are manifested in F1-score gains, balancing precision and recall, where significant increases have been noted, most especially for the Longitudinal class, which rose from 92.92% into 99.75%, reaching a perfect 100% for both Transverse and Crocodile classes. Overall accuracy adheres to the substantial gain wherein ResNet50-SE hit 99.88% accuracy at 93.25% in ResNet50. This underlines RESNET50-SE as truly more reliable and precise in classifying all instances correctly.

A comparison evaluation of model accuracy and confidence levels between ResNet50 and its improved equivalent, ResNet50-SE, demonstrates significant performance differences, as illustrated in Figures 3 and 4. Figure 3 illustrates the baseline performance of the ResNet50 model, which, although proficient in reliably detecting specific crack types, demonstrates significant discrepancies. This standard model often produces false positives, misclassifying longitudinal cracks as crocodile cracks with a confidence level of 55.34% and wrongly assessing crack width with a confidence of 65.49%. These findings highlight the intrinsic limits of ResNet50, especially its difficulties in distinguishing between crack types in diverse or intricate image situations. These limits indicate that the model's feature extraction and recognition skills are not robust, hence reducing its reliability in real applications related to structural health monitoring.

In contrast, Figure 4 highlights the impact of including the SE mechanism into ResNet50, leading to the ResNet50-SE model. This alteration significantly enhances categorization precision and confidence across all crack categories. The SE mechanism enhances the model's capacity to dynamically adjust features, enabling it to concentrate more accurately on the specific characteristics of each crack type. This recalibration results in nearly perfect classification accuracy, with confidence levels attaining 100% for various crack types. This heightened confidence highlights the SE mechanism's crucial function in enhancing both sensitivity and specificity, enabling the model to effectively distinguish across fracture types, a task that the original model found challenging.

This analysis indicates that incorporating the SE module markedly improves the model's reliability, resilience, and overall accuracy. This breakthrough is vital for the engineering sector, since it enables precise and reproducible crack categorization, a fundamental necessity for structural health monitoring and proactive maintenance planning. The ResNet50-SE model provides engineers and practitioners with a powerful instrument for reliable fracture detection, efficiently fulfilling essential infrastructure monitoring requirements. This advancement enhances the viability of automated crack identification and lays the groundwork for scaled applications in civil engineering, where reliable and accurate monitoring systems are essential for prompt repair and enduring infrastructure resilience.

Table 2. Comparison of ResNet50 and ResNet50-SE performance in classifying multiple crack type.

Crack Type	Metric	ResNet50 (%)	ResNet50-SE (%)	Performance improvement (%)
Longitudinal	Precision	87.95	99.50	**11.55**
	Recall	98.50	100.00	1.50
	F1-Score	92.92	99.75	**6.83**
Transverse	Precision	96.22	100.00	3.78
	Recall	89.00	100.00	**11.00**
	F1-Score	92.47	100.00	**7.53**
Crocodile	Precision	90.05	100.00	**9.95**
	Recall	86.00	100.00	**14.00**
	F1-Score	87.98	100.00	**12.02**
Non-crack	Precision	99.50	100.00	0.50
	Recall	99.50	99.50	-
	F1-Score	99.50	99.75	0.25
Overall Accuracy		93.25	**99.88**	**6.63**

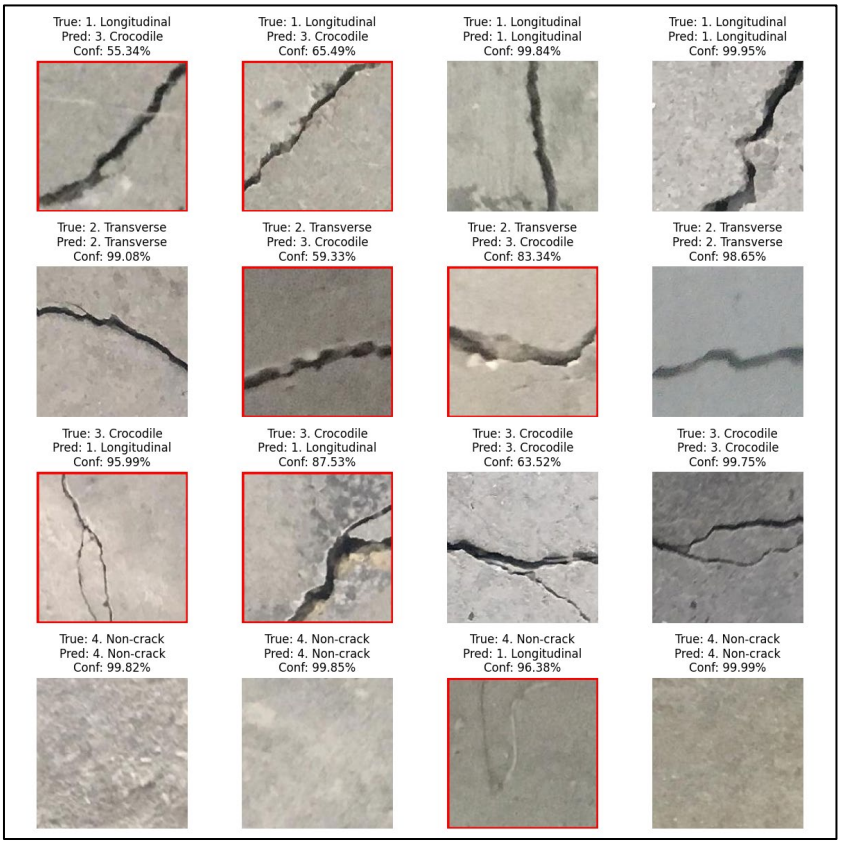

Figure 3. Performance of the ResNet50 model in crack classification.

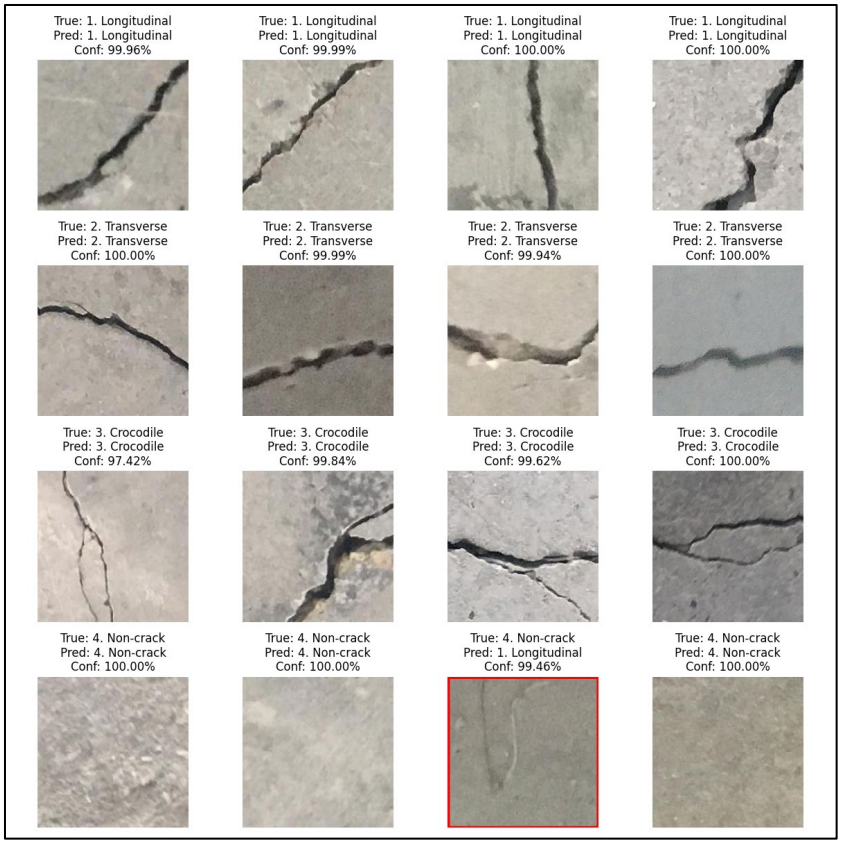

Figure 4. Enhanced crack classification performance of the ResNet50-SE Model.

4.3 Comparison with Existing Literature

To contextualize the effectiveness of our ResNet50-SE model, we compare its results against prior studies on crack classification. Table 3 summarizes the performance of different crack classification models reported in previous research.

Several studies, such as Bandi *et al.* [18] and Mazni *et al.* [19], have demonstrated high accuracy in binary crack classification, with MobileNetV2 achieving 99.87% accuracy. However, these approaches primarily focused on distinguishing between cracked and non-cracked surfaces, making them unsuitable for multi-class crack classification.

In contrast, multi-class crack classification has been addressed in recent works. Islam *et al.* [20] used AlexNet on the CCIC dataset (40,000 images), achieving 99.90% accuracy, 99.92% precision, and 99.80% recall. Their study demonstrated high classification accuracy; however, it relied on a specific dataset, making it less generalizable across various environmental conditions. Wibowo *et al.* [21] and Anggara *et al.* [22] employed ResNet50 with kNN and SVM classifiers, achieving 87%–89% accuracy. However, these models struggled with feature extraction under complex backgrounds and lighting variations, leading to lower precision and recall values. In contrast, the proposed ResNet50-SE model significantly outperforms these previous studies, achieving a 99.88% accuracy rate, with 100% precision and recall for three out of four crack types. This improvement is largely due to the SE blocks, which dynamically recalibrate feature maps and enhance model adaptability. Furthermore, previous works such as Yang *et al.* [12] attempted to mitigate noisy environments using Transfer Learning, achieving 97.5% accuracy. However, our model demonstrates greater robustness in handling real-world variations in crack images, as reflected in higher precision, recall, and F1-score values. Table 3 provides a comparative summary of the accuracy, precision, recall, and dataset sizes across different crack classification models, highlighting the superiority of the proposed ResNet50-SE framework.

Table 3. Comparison of ResNet50-SE with existing literature.

Study	Model Used	Crack Type	Dataset Size	Accuracy (%)	Precision (%)	Recall (%)	Key limitations
Bandi *et al.* [18]	ResNet50	Binary (crack vs. no crack)	56,000 (SDNET2018)	High (Exact not reported)	Not reported	Not reported	Affected by shadows
Mazni *et al.* [19]	MobileNetV2	Binary (crack vs. no crack)	11,435 images (Mendeley crack dataset)	99.87	99.74	100	Only binary classification
Islam *et al.* [20]	AlexNet	Multi-class (CCIC dataset)	40,000 images	99.9	99.92	99.8	Dataset dependence
Wibowo *et al.* [21]	ResNet50 + kNN	Multi-class (severity: light, medium, severe)	2,516 images	87.8	89.2	86.5	Small dataset, collected via web scraping
Anggara *et al.* [22]	SVM, RF, DT	Multi-class (crack orientation)	40,000 images	87.0 (SVM best)	87.0 (SVM best)	92.0 (SVM best)	Struggles with lighting variations
Yang *et al.* [12]	CNN + Transfer Learning	Multi-class	Three datasets (CCIC, SDNET2018, BCD) - Combined total ~19,816 images	98.54 (99.83% CCIC, 97.07% SDNET, 99.72% BCD)	Not reported	Not reported	Affected by lighting conditions
This Study (ResNet50-SE)	**ResNet50 + SE Block**	Multi-class (longitudinal, transverse, crocodile, non-crack)	7,147 images (4 classes, multiple sources)	**99.88**	**100% (3 classes)**	**100% (3 classes)**	No major limitations, robust across conditions

5. CONCLUSION

The incorporation of SE blocks in the ResNet50 model markedly enhances the efficacy of multi-class crack classification. The overall accuracy rises to 99.88% with ResNet50-SE in this research, compared to 93.25% in the original ResNet50 model. This enhancement illustrates the model's efficacy in accurately identifying and distinguishing between crack types. Compact and efficient architecture makes it well-suited for real-world applications in structural health monitoring, ensuring timely crack detection to support effective infrastructure maintenance.

ACKNOWLEDGMENT AND FUNDING

The authors would like to extend their hearty thanks for the financial assistance given for this research. This project was supported by UTM HIR Q.J13000.245108G87 and MOHE FRGS R.J130000.78085F400 research grants. Furthermore, the

authors are thankful and grateful for the support given by Universiti Teknologi MARA Cawangan Johor (UiTM Johor) via Grant 600-TNCPI 5/360/DDN (01) (001/2021).

DECLARATION OF CONFLICTING INTERESTS

The authors declare no competing interest in the research and publication of this article.

REFERENCES

[1] D. Kang, S. S. Benipal, D. L. Gopal and Y.-J. Cha, Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning, *Automation in Construction*, 118, 2020, 103291.
[2] B. Kim and S. Cho, Automated vision-based detection of cracks on concrete surfaces using a deep learning technique, *Sensors*, 18(10), 2018, 3452.
[3] F. Fang, L. Li, Y. Gu, H. Zhu and J. H. Lim, A novel hybrid approach for crack detection, *Pattern Recognition*, 107, 2020, 107474.
[4] S. Dorafshan, R. J. Thomas and M. Maguire, Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete, *Construction and Building Materials*, 186, 2018, 1031-1045.
[5] Y. Liu, J. K. W. Yeoh and D. K. H. Chua, Deep learning–based enhancement of motion blurred UAV concrete crack images, *Journal of Computing in Civil Engineering*, 34(5), 2020, 04020028.
[6] M. Mazni, A. R. Husain, M. I. Shapiai and I. S. Ibrahim, Crack recognition in masonry structures : CNN models with limited data sets, *ELEKTRIKA: Journal of Electrical Engineering*, 23(1), 2024, 133-140.
[7] D. Liang, X. F. Zhou, S. Wang and C. J. Liu, Research on concrete cracks recognition based on dual convolutional neural network, *KSCE Journal of Civil Engineering*, 23(7), 2019, 3066-3074.
[8] H. Xu, X. Su, H. Xu and H. Li, Autonomous bridge crack detection using deep convolutional neural networks, *Proceedings of the 3rd International Conference on Computer Engineering, Information Science & Application Technology*, 2019.
[9] S. Li and L. Sun, Detectability of bridge-structural damage based on fiber-optic sensing through deep-convolutional neural networks, *Journal of Bridge Engineering*, 25(4), 2020, 04020012.
[10] E. McLaughlin, N. Charron and S. Narasimhan, Automated defect quantification in concrete bridges using robotics and deep learning, *Journal of Computing in Civil Engineering*, 34(5), 2020, 04020029.
[11] T. S. Tran, V. P. Tran, H. J. Lee, J. M. Flores and V. P. Le, A two-step sequential automated crack detection and severity classification process for asphalt pavements, *International Journal of Pavement Engineering*, 23(6), 2022, 2019-2033.
[12] Q. Yang, W. Shi, J. Chen and W. Lin, Deep convolution neural network-based transfer learning method for civil infrastructure crack detection, *Automation in Construction*, 116, 2020, 103199.
[13] T. U. Ahmed, M. Shahadat Hossain, M. J. Alam and K. Andersson, An integrated CNN-RNN framework to assess road crack, *22nd International Conference on Computer and Information Technology (ICCIT 2019)*, Dhaka, Bangladesh, 2019, 18-20.
[14] L. Ali, F. Alnajjar, W. Khan, M. A. Serhani and H. A. Jassmi, Bibliometric analysis and review of deep learning-based crack detection literature published between 2010 and 2022, *Buildings*, 12(4), 432.
[15] 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, *Applications of Modelling and Simulation*, 8(1), 2024, 1-25.
[16] M. A. M. Khan, S. H. Kee, A. S. K. Pathan and A. Al Nahid, Image processing techniques for concrete crack detection: A scientometrics literature review, *Remote Sensing*, 15(9), 2023, 1-46.
[17] A. Reghukumar and L. J. Anbarasi, Crack detection in concrete structures using deep learning, *Lecture Notes in Electrical Engineering*, 711, 2021, 211-219.
[18] M. R. Bandi, L. N. Pasupuleti, T. Das and S. Guchhait, Deep learning based damage detection of concrete structures, *Asian Journal of Civil Engineering*, 25, 2024,5197-5204.
[19] M. Mazni, A. R. Husain, M. I. Shapiai, I. S. Ibrahim, D. W. Anggara and R. Zulkifli, An investigation into real-time surface crack classification and measurement for structural health monitoring using transfer learning convolutional neural networks and Otsu method, *Alexandria Engineering Journal*, 92, 2024, 310-320.
[20] M. M. Islam, M. B. Hossain, M. N. Akhtar, M. A. Moni and K. F. Hasan, CNN based on transfer learning models using data augmentation and transformation for detection of concrete crack, *Algorithms*, 15(8), 2022, 287.
[21] A. P. Wibowo, A. Adha, I. F. Kurniawan and I. Laory, Wall crack multiclass classification: expertise-based dataset construction and learning algorithms performance comparison, *Buildings*, 12(12), 2022, 2135.
[22] D. W. Anggara *et al.*, Concrete crack detection, orientation and measurement using a wall climbing robot, *Applications of Modelling and Simulation*, 8(1), 2024, 272-282.
[23] M. Mazni, A. R. Husain, M. I. Shapiai, I. S. Ibrahim, R. Zulkifli and D. W. Anggara, Real-time crack classification with wall-climbing robot using MobileNetV2, *Communications in Computer and Information Series: Methods and Applications for Modeling and Simulation of Complex Systems*, 1911, 2024, 319-328.
[24] R. Zulkifli, A. R. Husain, I. S. Ibrahim, M. Mazni and N. H. A. M. Fauzan, Analysis of the hybrid adhesion mechanism of the wall climbing robot, *Lecture Notes in Electrical Engineering: Control, Instrumentation and Mechatronics: Theory and Practice*, 921, 2022, 155-169.
[25] F. Sukarman, M. A. Mat Shah, M. Mazni, N. Abdul Malik and M. G. Mohd Hamami, Integrated concurrent programmable logic controller (PLC) based IR 4.0 training kit, *The 9th International Innovation, Invention and Design Competition 2020*, 2020, 360-364.

[26] F. Özgenel and A. Gönenç 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*, 2018.
[27] E. Bianchi and M. Hebdon, *Development of Extendable Open-Source Structural Inspection Datasets*, 36, 2022.
[28] Y. Liu, J. Yao, X. Lu, R. Xie and L. Li, DeepCrack: A deep hierarchical feature learning architecture for crack segmentation, *Neurocomputing*, 338, 2019, 139-153.