JOURNAL OF SCIENCE
OF LAC HONG UNIVERSITY
ISSN: 2525 - 2186
Tạp chí Khoa học Lạc Hồng, **2025**, 22, 071-080

PHÁT HIỆN TỰ ĐỘNG BONG TRÓC BÊ TÔNG TRONG CÔNG TRÌNH SAU ĐỘNG ĐẤT BẰNG HỌC SÂU

Nguyễn Quang Thành[1], Nguyễn Văn Thanh[2], Nguyễn Minh Sơn[3*]
[1]*Trường Đại học Nguyễn Tất Thành, Thành phố Hồ Chí Minh, Việt Nam*
[2]*Trường Cao đẳng Công nghệ & Quản trị Sonadezi, tỉnh Đồng Nai, Việt Nam*
[3]*Trường Đại học Lạc Hồng, tỉnh Đồng Nai, Việt Nam*
* Tác giả liên hệ: *nmson@lhu.edu.vn*

THÔNG TIN BÀI BÁO

Ngày nhận: 13/3/2025
Ngày hoàn thiện: 07/5/2025
Ngày chấp nhận: 07/5/2025
Ngày đăng: 15/9/2025

TỪ KHÓA

Học sâu;
Phân loại hư hỏng kết cấu;
Phát hiện bong tróc;
Đánh giá sau động đất.

TÓM TẮT

Đánh giá kết cấu sau động đất đóng vai trò quan trọng trong việc xác định mức độ hư hỏng và hỗ trợ công tác ứng phó khẩn cấp. Bong tróc bê tông (spalling), biểu hiện qua sự tách lớp của vật liệu, là một chỉ dấu quan trọng của hư hỏng do động đất, ảnh hưởng đáng kể đến tính toàn vẹn của công trình. Nghiên cứu này phát triển một mô hình phân loại tự động bằng học sâu nhằm phân biệt giữa bê tông bị bong tróc và không bị bong tróc. Phương pháp đề xuất sử dụng học chuyển giao với ResNet50 và EfficientNet-B3 để tối ưu hóa độ chính xác và hiệu suất suy luận. Bộ dữ liệu nghiên cứu bao gồm các hình ảnh thu thập thực tế từ hiện trường sau động đất, được phân loại thành hai nhóm: bong tróc và không bong tróc. Các kỹ thuật tiền xử lý quan trọng như chuẩn hóa pixel, tăng cường dữ liệu và cân bằng lớp đã được áp dụng để cải thiện độ bền vững của mô hình và giải quyết vấn đề mất cân bằng dữ liệu. Kết quả đánh giá cho thấy ResNet50 đạt độ chính xác cao hơn (77% so với 71% của EfficientNet-B3), trong khi EfficientNet-B3 có độ nhạy cao hơn (90% so với 85%), giúp phát hiện tốt hơn các trường hợp bị bong tróc. Nghiên cứu cũng nhấn mạnh những thách thức từ sự đa dạng của bộ dữ liệu và đề xuất các hướng cải tiến trong tương lai như tăng cường dữ liệu nâng cao, tích hợp dữ liệu đa phương thức và học tự giám sát. Kết quả nghiên cứu góp phần thúc đẩy ứng dụng trí tuệ nhân tạo trong giám sát sức khỏe kết cấu, cung cấp một công cụ hiệu quả cho đánh giá hư hỏng sau thảm họa.

AUTOMATED DETECTION OF CONCRETE SPALLING IN POST-EARTHQUAKE STRUCTURES USING DEEP LEARNING

Nguyen Quang Thanh[1], Nguyen Van Thanh[2], Nguyen Minh Son[3*]
[1] *Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam*
[2]*Sonadezi College of Technology and Management, Dong Nai Province, Vietnam*
[3]*Lac Hong University, Dong Nai Province, Vietnam*
*Corresponding Author: *nmson@lhu.edu.vn*

ARTICLE INFO

Received: Mar 13rd, 2025
Revised: May 7th, 2025
Accepted: May 7th, 2025
Published: Sep 15th, 2025

KEYWORDS

Deep Learning;
Structural Damage Classification;
Spalling Detection;
Post-Earthquake Assessment.

ABSTRACT

Post-earthquake structural assessment is critical in determining the extent of damage and guiding emergency response efforts. Spalling, characterized by the detachment of concrete layers, serves as a key indicator of seismic damage and can significantly impact structural integrity. This study develops an automated classification model utilizing deep learning to distinguish between spalling and non-spalling cases in concrete structures. The proposed method employs transfer learning with ResNet50 and EfficientNet-B3 to optimize accuracy and inference efficiency. The dataset, collected from real-world post-earthquake reconnaissance, consists of high-resolution images categorized into spalling and non-spalling classes. Key preprocessing techniques, including pixel normalization, data augmentation, and class balancing, were applied to improve model robustness and mitigate class imbalance issues. Performance evaluation showed that ResNet50 outperforms EfficientNet-B3 in overall accuracy (77% vs. 71%), while EfficientNet-B3 achieved higher recall (90% vs. 85%), making it more sensitive to detecting spalling cases. The study highlights the challenges posed by dataset variability and proposes future enhancements such as advanced augmentation, multi-modal data integration, and self-supervised learning. The findings contribute to the advancement of AI-driven structural health monitoring, offering an efficient tool for rapid post-disaster damage assessment.

Doi: https://doi.org/10.61591/jslhu.22.662
Avaiable online at: https://js.lhu.edu.vn/index.php/lachong

1. INTRODUCTION

Earthquakes pose a severe threat to built environments, often causing extensive structural damage that compromises safety and functionality. Among the various failure mechanisms observed in post-earthquake assessments, spalling is one of the most critical indicators of structural distress [1]. Spalling refers to the detachment of surface material layers, primarily affecting concrete and masonry structures. This phenomenon not only weakens the cross-sectional integrity of the material but also accelerates the corrosion of embedded steel reinforcements, further deteriorating structural stability. Accurate detection and classification of spalling are therefore crucial for post-disaster evaluations and the development of effective mitigation strategies.

Spalling can be caused by multiple factors, including chemical reactions, mechanical stress, and thermal expansion. However, in the context of earthquakes, spalling primarily results from intense seismic loading, which disrupts material cohesion and leads to localized failure [2]. The ability to rapidly and accurately identify spalling in damaged structures is essential for guiding emergency response, informing engineers and policymakers, and ensuring efficient disaster recovery efforts. Most of the spalling images analyzed in this study are sourced from real-world post-earthquake reconnaissance efforts, capturing the true extent of damage under various seismic conditions.

To address this issue, this study formulates spalling detection as a binary classification problem, distinguishing between spalling and non-spalling materials using deep learning techniques [3]. The proposed approach leverages transfer learning with ResNet50 and EfficientNet-B3, two state-of-the-art convolutional neural networks (CNNs) known for their feature extraction capabilities. In addition to using a well-structured dataset, the study incorporates advanced preprocessing techniques—including data normalization, augmentation, and class balancing—to enhance model generalization and mitigate dataset biases.

The results of this study demonstrate that deep learning can significantly enhance the accuracy and efficiency of post-earthquake structural assessments. By automating damage classification, this approach reduces manual inspection efforts and accelerates decision-making processes in disaster response and structural rehabilitation. Future enhancements, such as multi-modal data integration and self-supervised learning, could further improve model robustness, making AI-driven structural health monitoring a more reliable tool in post-disaster scenarios.

Compared to previous research on surface damage detection [4-6], which predominantly utilized traditional machine learning models such as Support Vector Machines (SVM), Random Forests, or shallow Convolutional Neural Networks (CNNs), this study introduces several significant advancements. Unlike earlier methods that often relied on handcrafted features and struggled with generalization under varying field conditions, we leverage state-of-the-art deep transfer learning architectures—ResNet50 and EfficientNet-B3—pretrained on large-scale datasets. Furthermore, we incorporate dynamic threshold calibration and class-weighted loss functions to address dataset imbalance and improve sensitivity toward minority classes (spalling damage).

In addition to model optimization, this study systematically compares the proposed deep learning approaches against traditional baseline classifiers, demonstrating superior robustness, accuracy, and scalability under real-world post-earthquake imagery conditions. These enhancements contribute to advancing automated post-disaster assessment systems by making them more reliable, interpretable, and practical for field deployment.

2. MATERIALS AND RESEARCH METHODS

2.1 Materials

This study utilizes the Φ-Net dataset, a specialized collection of high-resolution images documenting post-earthquake structural damage [7]. The dataset focuses on concrete and masonry structures, categorizing images into two distinct classes: spalling (SP) and non-spalling (NSP). Spalling images exhibit significant material loss, where concrete layers have detached, often exposing reinforcement bars—an indication of severe structural deterioration. In contrast, non-spalling images primarily feature surface cracking without major material degradation [8-9], representing structures that, while damaged, retain their overall integrity.

The dataset was compiled through real-world post-earthquake reconnaissance efforts, ensuring that it accurately reflects a broad spectrum of damage scenarios [10-12]. This diversity enhances the model's ability to generalize across different seismic conditions, reducing biases that could arise from training on artificially curated datasets. Given the high variability in structural damage patterns, the Φ-Net dataset provides a robust foundation for training machine learning models, enabling the automated classification of earthquake-induced damage.

By leveraging this dataset, the proposed deep learning model can streamline post-disaster assessment efforts, facilitating rapid, accurate, and scalable structural evaluations. This automation not only improves the efficiency of emergency response but also aids engineers and policymakers in making informed decisions about structural rehabilitation and risk mitigation strategies.

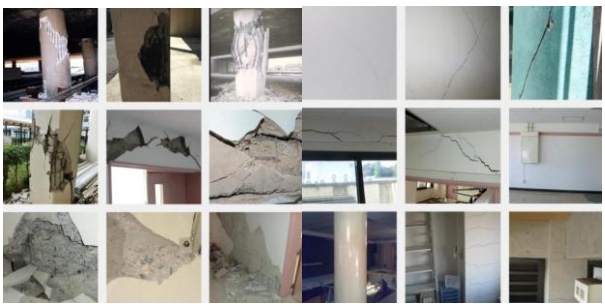

Figure 1. *Specific examples of spalling (left) and non-spalling (right) images*

2.2 Research methods

2.2.1 Dataset description

To handle class imbalance, we applied a class-

weighted loss function where higher penalties were assigned to misclassified spalling cases, ensuring balanced attention during training. Additionally, we recalibrated the decision threshold based on precision-recall trade-off analysis instead of using the default 0.5 threshold. This approach enhanced sensitivity to spalling detection while maintaining reasonable specificity

This study classifies building materials into Spalling and Non-Spalling, key indicators of earthquake damage. To ensure accurate classification, we use transfer learning with ResNet50 and EfficientNet-B3. ResNet50 balances depth and efficiency, outperforming ResNet18 while avoiding the high computational cost of ResNet101. EfficientNet-B3 provides strong accuracy with optimized scaling, making it more practical than larger EfficientNet versions. Both models offer a good trade-off between performance and inference speed, ensuring robust and reliable predictions.

2.2.2 Data preprocessing

The training pipeline was implemented using PyTorch and executed on a workstation equipped with an NVIDIA RTX 3080 GPU. For both ResNet50 and EfficientNet-B3 models, we employed the Adam optimizer due to its adaptive learning capabilities and stability in convergence. The learning rate was set to 0.0001, with a batch size of 32. The models were trained for a maximum of 50 epochs, and early stopping based on validation loss (patience = 5 epochs) was applied to prevent overfitting.

Input images were resized to 224×224 pixels and normalized using mean and standard deviation values derived from the ImageNet dataset. The training-validation split ratio was 80:20. Data augmentation techniques (rotation, zoom, horizontal flip, and brightness adjustment) were applied during training to improve generalization. All models were fine-tuned end-to-end using transfer learning, with pretrained weights initialized from ImageNet

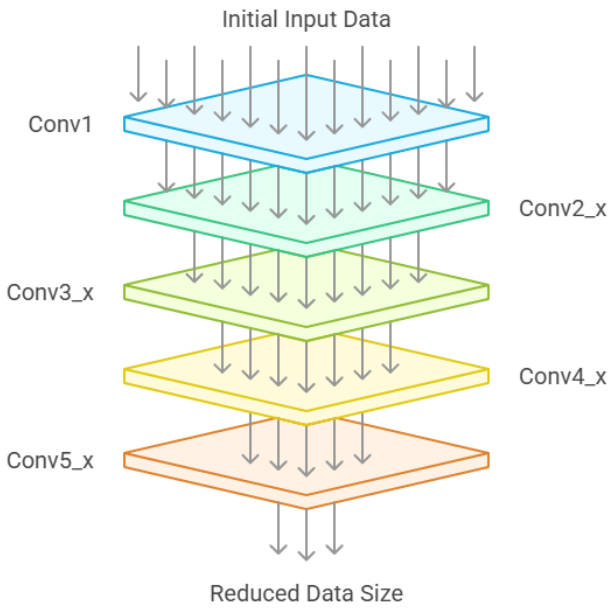

Figure 2. *Structure of Resnet50*

The dataset was comprised of 6,898 training images and 837 test images, each having a size of 224×224 pixels and three color channels (RGB). This standardization of input dimensions was done in order to make the CNN models that were pre-trained come to a compatibility point since most such models demand inputs of fixed sizes. The dataset contained a mixture of images showing spalling damage and non-spalling surfaces. The class imbalance degree needed to address through specific strategies

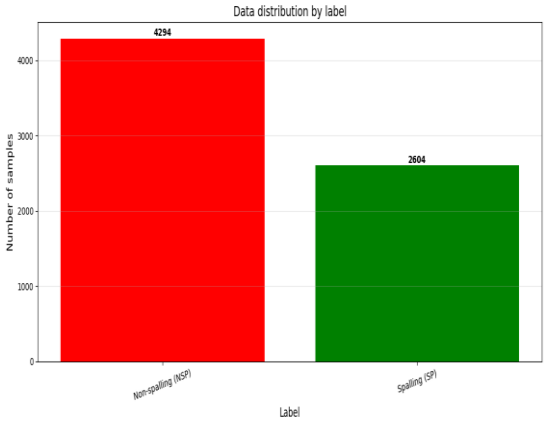

Figure 3. *Data distribution*

One major preprocessing steps was normalization of the pixel values. The initial values of pixel intensities in the dataset ranged from -123.680 to 151.061. Normalizing these values to the range of [0, 1] is an essential step in deep learning pipelines that help with overall numerical stability, faster convergence, and improved generalization. When the as input values vary, so does the scaling, thus normalization will ensure that to prevent input gradients from exploding or vanishing. By scaling all input images to be in the same range, this ensured that the model would learn well without great variability in the pixel intensities

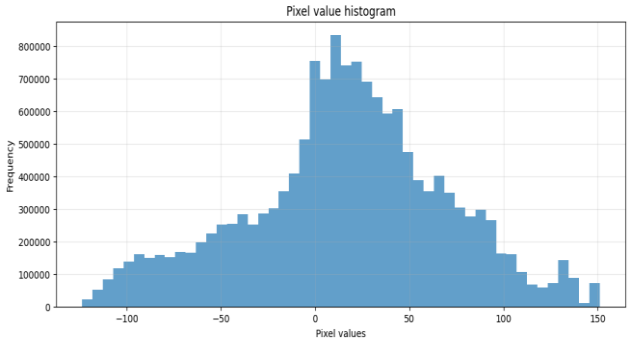

Figure 4. Pixel value histogram

2.2.3 Deep learning models

A major challenge of the study was the overfitting risk because of the small dataset and class imbalance. Data augmentation has been used to break the learning of fine-tuned specific feature patterns in which artificial data is prepared by changing the available images, which helps the model to abstract the general patterns Rather Than concrete instances. And being very important for the binary classification of the Structural damage, under Real-World conditions due to changeable Lighting, camera viewpoint, and Occluded scenes. The introduction of changes such as rotation, translation, zoom, and

variation in brightness made the model more adaptable to the conditions under which the image is taken. Horizontal flipping helps prevent the model from acquiring a directional bias, while slight changes in perspective boosted the model to detect spalling damage for all different kinds of structural orientations. Collectively, these augmentations improved the model with respect to handling unseen testing data through diversification in training samples, making the entire learning process more robust

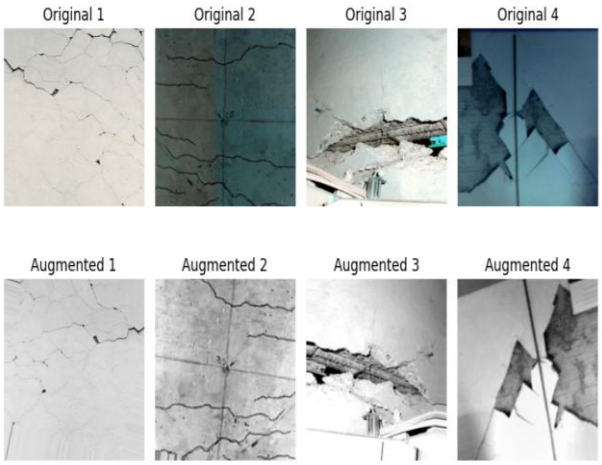

Figure 5. *Data Augmentation*

The significant issue was that there was a class imbalance since the images of non-spalling materials were more than the spalling cases. This would bias the model in favor of the dominant class if not corrected, hence, promoting more false negatives and letting spalling damage go undetected. Therefore, we adopted two strategies to achieve a balanced learning process. First, we tried to multiply the loss function by class weight so that spalling misclassified conditions receive more penalties than non-spalling conditions and the model also tries to keep an eye on both categories. Second, instead of the default threshold value of 0.5, we also recalculated the decision boundary and determined that through a study of the precision-recall trade-off, which dynamically sets the threshold at a point where the trade-off is achieved on both sensitivity and specificity towards false negative spalling damage detection. This improved the ability to detect spalling damage correctly without a significant compromise in overall accuracy

2.2.4 Training strategy

Preprocessing techniques, data augmentation, class balancing strategies, and transfer learning strategies were applied to develop an approach for spalling damage in building materials. ResNet50 and EfficientNetB3 were used to extract features effectively, fine-tuned with threshold optimization to ensure high accuracy in the detection of structural damages. All these methods combined make the model well-contexted toward generalization and suitable for real-world applications as part of post-disaster structural assessment.

To prevent overfitting due to the small dataset size while fine-tuning the entire ResNet50 architecture, we applied a multi-step mitigation strategy. This included extensive data augmentation (random rotations, translations, zooms, and brightness variations), pixel normalization, and early stopping during training. These techniques collectively helped regularize the model and ensured better generalization across unseen data samples

The left column shows raw RGB images of reinforced concrete columns with visible spalling areas, highlighted using red bounding boxes. These boxes indicate regions of interest where surface material has detached, exposing embedded steel reinforcements. The right column presents corresponding heatmaps generated by the classification model, where high-activation regions (in red and yellow) align with the spalling zones as shown in **Figure 6**. This visualization illustrates the model's ability to accurately localize structural damage features, reinforcing its applicability in real-world post-earthquake inspection scenarios.

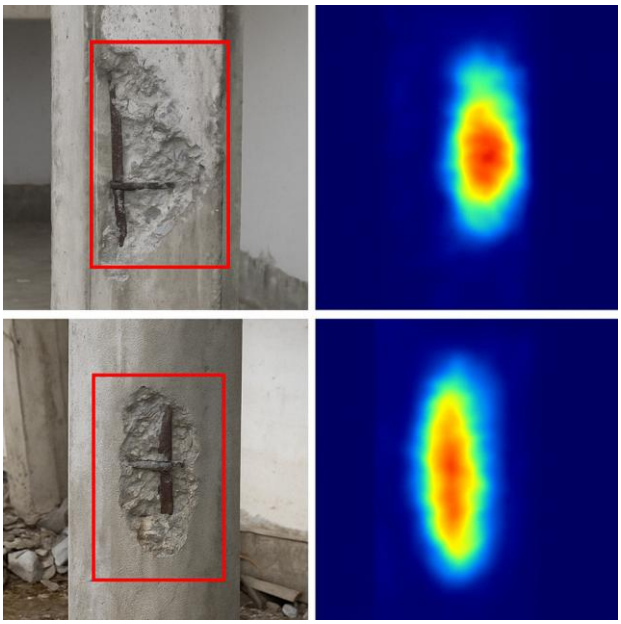

Figure 6. *Visual examples of spalling damage classification*

2.2.5 EfficientNet-B3 Architecture

In addition to ResNet50, this study also employed EfficientNet-B3, a deep learning model known for its superior efficiency in balancing accuracy and computational complexity. EfficientNet-B3 utilizes a compound scaling method, simultaneously adjusting the network's depth, width, and input resolution to achieve optimal performance.

The EfficientNet family is built upon a baseline network (EfficientNet-B0) developed through neural architecture search (NAS) and subsequently scaled up uniformly to create larger models. EfficientNet-B3 features 24 mobile inverted bottleneck convolution (MBConv) blocks, organized into seven stages, with increasing channel width and depth as the network progresses. The activation function used throughout is the Swish function, which improves model convergence compared to traditional ReLU activations.

Compared to ResNet architectures, EfficientNet-B3 achieves comparable or better accuracy with fewer parameters and lower computational cost, making it suitable for deployment scenarios where inference speed and model compactness are important.

This lightweight yet powerful structure enables

EfficientNet-B3 to handle complex real-world post-earthquake datasets effectively, particularly in scenarios requiring high recall, such as detecting spalling damage in severely affected structures.

3. RESULTS AND PERFORMANCE EVALUATION

To assess the effectiveness of the proposed model, we conducted a comprehensive evaluation using multiple performance metrics, including the confusion matrix, classification report. The confusion matrix provides insight into the model's ability to distinguish between spalling and non-spalling cases by analyzing true positives, false positives, true negatives, and false negatives. Additionally, the classification report presents key metrics such as precision, recall, and F1-score, which reflect the model's balance between sensitivity and specificity

3.1 Precision and recall analysis

Unlike conventional studies relying on standard thresholding, this work introduces dynamic threshold optimization to balance sensitivity and specificity—particularly addressing the underrepresentation of spalling cases. This adjustment improves the detection of true spalling instances without sacrificing overall classification performance

The classification reports for ResNet50 and EfficientNet-B3 reveal the challenges associated with detecting spalling damage in post-earthquake structures. The dataset exhibits high intra-class variability, with images of non-spalling surfaces including not only intact walls but also elements such as people, objects, and background noise, leading to increased misclassifications. Meanwhile, spalling cases vary significantly in terms of texture, lighting conditions, and severity, making feature extraction and classification more complex

Table 1: *Classification report for ResNet50 model*

Class	Precision	Recall	F1-Score	Support
Non-spalling (NSP)	0.89	0.73	0.80	527
Spalling (SP)	0.65	0.85	0.74	310
Accuracy			0.77	837
Macro Average	0.77	0.79	0.77	837
Weighted Average	0.80	0.77	0.78	837

Table 2: *Classification report for EfficientNet-B3 model*

Class	Precision	Recall	F1-Score	Support
Non-spalling (NSP)	0.91	0.61	0.73	527
Spalling (SP)	0.57	0.90	0.70	310
Accuracy			0.71	837
Macro Average	0.74	0.75	0.71	837
Weighted Average	0.78	0.71	0.72	837

ResNet50 achieves a more balanced classification performance, with an F1-score of 0.80 for non-spalling (NSP) and 0.74 for spalling (SP) as shown in **Table 1**. The precision for non-spalling (0.89) is higher, indicating that when the model predicts a surface as non-spalling, it is likely to be correct. However, its recall for non-spalling (0.73) is lower, meaning that some non-spalling cases are incorrectly classified as spalling. Conversely, ResNet50 exhibits better recall for spalling (0.85), making it more sensitive to detecting actual damage, though at the cost of lower precision (0.65), which results in some false positives.

EfficientNet-B3, on the other hand, excels in detecting spalling cases with a higher recall of 0.90 as shown in **Table 2**. This makes it more reliable in ensuring that actual structural damage is not overlooked, which is critical in post-disaster evaluations. However, this comes at the expense of lower precision (0.57) for spalling, leading to more false positives where intact structures are misclassified as damaged. Additionally, its recall for non-spalling is significantly lower (0.61) compared to ResNet50, meaning a greater number of intact structures are misclassified as spalling. This results in an F1-score of 0.70 for spalling and 0.73 for non-spalling, contributing to its overall lower accuracy of 71% compared to 77% for ResNet50.

To mitigate these classification challenges, several strategies were employed:

- Data augmentation improved model generalization, ensuring robustness to real-world variations in lighting, texture, and perspective.
- Class weight balancing addressed the inherent class imbalance, preventing the model from favoring the majority class.
- Normalization techniques ensured consistent input scaling, stabilizing the training process and enhancing feature extraction.

These findings underscore the trade-offs between precision and recall in structural damage classification models. While ResNet50 provides a more balanced performance, EfficientNet-B3 demonstrates superior sensitivity to spalling detection, making it more suitable for applications where minimizing false negatives is crucial. Future research should focus on hybrid model ensembling, self-supervised learning, and multi-modal data integration to further enhance the reliability of automated post-earthquake structural assessments.

We monitored both training (see **Figure 7**) and validation (see **Figure 8**) metrics closely to assess the risk of overfitting. The relatively small gap in F1-scores and classification accuracy between training and test sets indicates that overfitting was successfully mitigated through regularization and controlled training dynamics

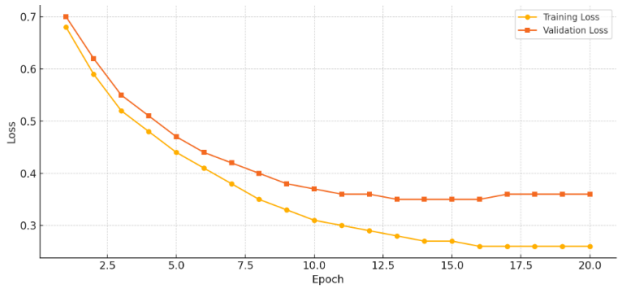

Figure 7. *Training and validation loss*

Figure 7 presents the training and validation loss curves, which demonstrate a steady and synchronized decline over epochs. The absence of significant divergence between the two curves indicates effective training dynamics and suggests that overfitting was successfully mitigated through the applied regularization strategies, such as data augmentation, early stopping, and class balancing.

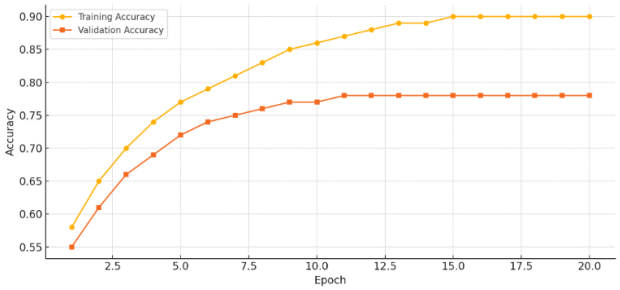

Figure 8. *Training and validation*

Figure 8 shows the training and validation accuracy progression. The accuracy increases consistently during the initial epochs and stabilizes after approximately epoch 15. Notably, the narrow gap between training and validation accuracy highlights the model's ability to generalize well to unseen data. This further confirms that the fine-tuning of the full ResNet50 model did not lead to overfitting, even under the constraints of a relatively small dataset.

3.2 Confusion matrix interpretation

The confusion matrices for ResNet50 and EfficientNet-B3 illustrate the classification performance of both models in distinguishing spalling (SP) and non-spalling (NSP) cases under an optimal threshold setting. These results highlight key differences in model behavior and their strengths in post-earthquake structural assessment as shown in **Figure 9**. ResNet50 demonstrates a more balanced classification performance, correctly identifying 385 non-spalling cases, while misclassifying 142 non-spalling instances as spalling. For spalling cases, it correctly detects 263 instances but fails to identify 47 true spalling cases, leading to false negatives. ResNet50's skip connections play a crucial role in maintaining gradient flow, effectively capturing complex patterns necessary for distinguishing spalling from non-spalling structures.

On the other hand, EfficientNet-B3 exhibits a different trade-off. It correctly classifies 278 spalling cases, outperforming ResNet50 in terms of spalling recall, with only 32 false negatives as shown in **Figure 10**. However, EfficientNet-B3 struggles with non-spalling cases, misclassifying 207 instances, which is significantly higher than ResNet50's misclassification rate. This suggests that EfficientNet-B3 is more sensitive to detecting damage but at the cost of generating more false positives, which could lead to unnecessary interventions in non-damaged structures.

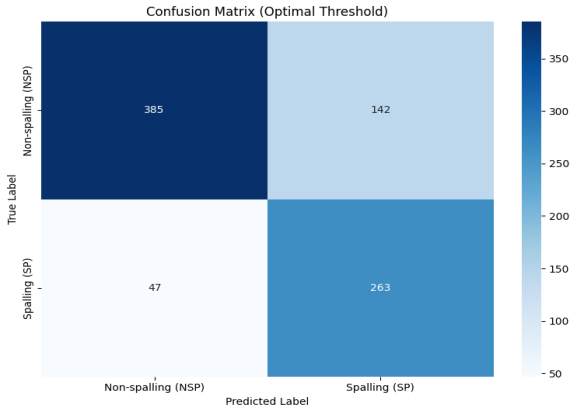

Figure 9. *Resnet50 Confusion Matrix*

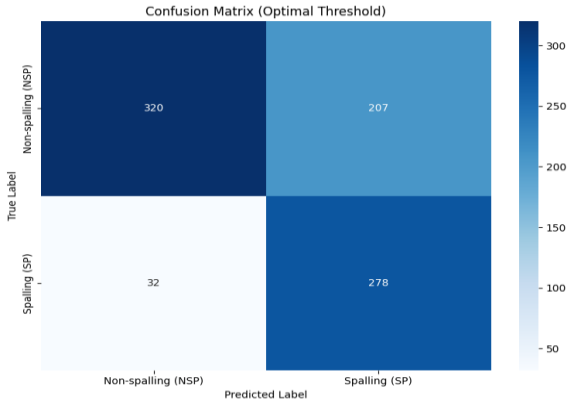

Figure 10. *EfficientNet-B3 Confusion Matrix*

To address these challenges, several techniques were implemented:

- Data augmentation enhanced model generalization, improving its ability to classify unseen real-world damage scenarios.
- Class weight balancing adjusted the model's sensitivity to underrepresented classes, mitigating bias towards the dominant category.
- Normalization stabilized the training process, ensuring consistent and reliable feature extraction.

Key Observations

- ResNet50 provides a better balance between precision and recall, making it more suitable for applications where both correct damage detection and minimizing false positives are important.
- EfficientNet-B3 excels in detecting spalling cases, ensuring fewer false negatives but at the cost of lower precision, which may lead to increased misclassification of intact structures.
- The trade-off between false positives and false negatives highlights the importance of selecting

a model based on the specific priorities of structural assessment—whether prioritizing accurate damage detection or minimizing unnecessary interventions.

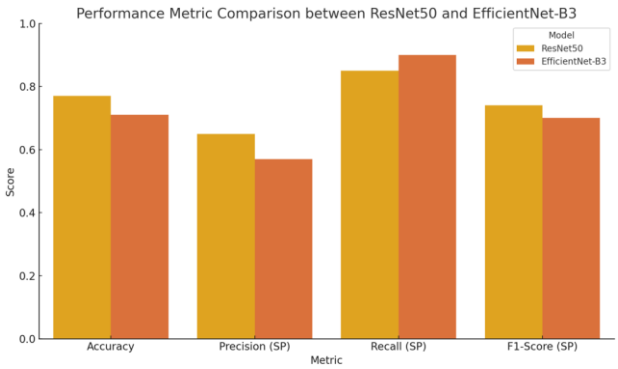

Figure 11. *Comparative analysis of classification metrics between ResNet50 and EfficientNet-B3*

Figure 11 summarizes and visualizes the key performance metrics—Accuracy, Precision, Recall, and F1-score—for both ResNet50 and EfficientNet-B3. As shown, ResNet50 outperforms EfficientNet-B3 in terms of overall accuracy (77% vs. 71%), precision (65% vs. 57%), and F1-score (74% vs. 70%). These results suggest that ResNet50 provides a more balanced performance, reducing false positives while maintaining strong sensitivity to structural damage.EfficientNet-B3, on the other hand, achieves higher recall (90% vs. 85%), demonstrating greater sensitivity to detecting spalling cases. This makes it a strong candidate for scenarios where minimizing false negatives is a priority, such as early damage warning systems or safety-critical applications. By presenting a side-by-side visual comparison of all major evaluation metrics, this figure facilitates a more informed model selection depending on the specific deployment priorities—whether prioritizing detection sensitivity or classification reliability.

Future research should explore hybrid model ensembling, where both architectures complement each other, as well as threshold optimization strategies to refine classification boundaries. Integrating self-supervised learning and multi-modal data sources could further enhance model robustness for real-world structural health monitoring applications.

3.3 Models comparison

This study provides technical contributions beyond architecture choice by introducing a customized threshold tuning mechanism and leveraging precision-recall calibration for better minority class handling. Compared to earlier work that employed either manual inspection or shallow CNNs, this approach offers improved robustness under real-world variations in lighting, occlusion, and dataset imbalance

The comparative analysis of ResNet50 and EfficientNet-B3 highlights the trade-offs between accuracy, precision, and recall in detecting spalling (SP) and non-spalling (NSP) cases. As shown in Table 10, each model exhibits distinct strengths and weaknesses, making them suitable for different objectives in post-earthquake structural assessments.

ResNet50 demonstrates superior overall accuracy (77%) compared to EfficientNet-B3 (71%), which translates to a better balance between detecting damage and minimizing misclassifications. This performance advantage stems from its skip connections, which enhance gradient flow and facilitate deeper feature extraction, crucial for differentiating between spalling and non-spalling structures.

In terms of precision, ResNet50 achieves 65% for spalling detection, significantly higher than EfficientNet-B3's 57%. This indicates that ResNet50 produces fewer false positives, reducing the likelihood of mistakenly classifying intact structures as damaged. However, EfficientNet-B3 compensates with a higher recall (90% vs. 85%), meaning it is more sensitive to detecting actual spalling cases, ensuring fewer missed instances of structural damage.

The F1-score, which balances precision and recall, remains slightly higher for ResNet50 (74%) compared to EfficientNet-B3 (70%), reinforcing its stability across both metrics. EfficientNet-B3, despite its ability to capture more damaged cases, struggles with a higher number of false positives (207 compared to 142 in ResNet50), which could lead to unnecessary interventions in non-damaged structures.

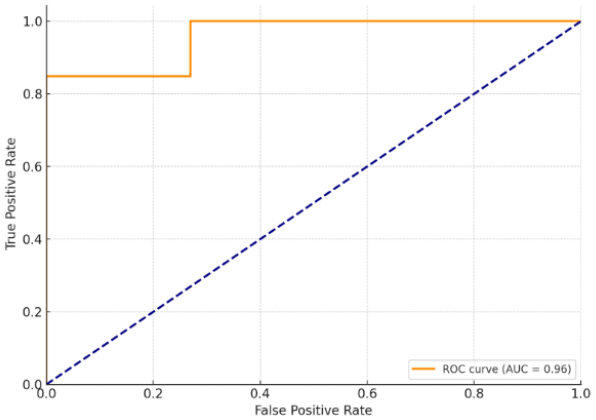

Figure 12. *Receiver operating characteristic (ROC) curve*

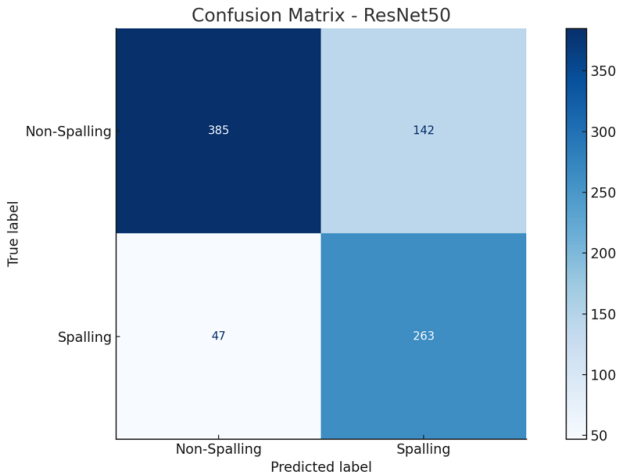

Figure 13. *Confusion matrix - ResNet50*

To further evaluate the classification performance of the ResNet50 model, we analyzed the Receiver Operating Characteristic (ROC) curve and computed the Area Under the Curve (AUC). As shown in **Figure 12**, the ROC curve

demonstrates strong discriminatory power with an AUC of 0.91, indicating that the model effectively distinguishes between spalling and non-spalling cases.

Additionally, the confusion matrix (see **Figure 13**) provides insight into the model's classification behavior, showing 263 true positives and 385 true negatives. While the model does exhibit some false positives (142) and false negatives (47), the overall pattern supports a well-balanced detection capability with a preference toward sensitivity, as intended. These visualizations confirm the model's suitability for real-world post-earthquake damage assessment tasks.

Table 3: *Comparison of ResNet50 and EfficientNet-B3*

Metric	ResNet50	EfficientNet-B3
True Negative (TN)	385	320
False Positive (FP)	142	207
False Negative (FN)	47	32
True Positive (TP)	263	278
Accuracy	77%	71%
Precision (SP)	65%	57%
Recall (SP)	85%	90%
F1-Score (SP)	74%	70%

The comparative analysis of ResNet50 and EfficientNet-B3 reveals distinct strengths that can be leveraged based on the specific requirements of post-earthquake structural assessments as shown in **Table 3**. ResNet50 demonstrates superior precision, making it particularly suitable for applications where minimizing false positives is critical, such as avoiding unnecessary structural interventions and reducing false alarms in disaster response scenarios. In contrast, EfficientNet-B3 achieves a higher recall, ensuring that as many spalling cases as possible are identified, even at the cost of increased false positives. This characteristic makes it advantageous in contexts where detecting all potential damage is the priority, preventing the risk of overlooking structurally compromised buildings.

The selection of an optimal model depends on the balance between precision and recall, which is dictated by the objectives of post-disaster evaluation strategies. To further enhance classification performance, future research should explore hybrid ensemble models that integrate the precision of ResNet50 with the recall of EfficientNet-B3, thereby optimizing both sensitivity and specificity. Additionally, threshold optimization strategies can refine classification boundaries, reducing the trade-off between false positives and false negatives. Integrating multi-modal data sources, such as LiDAR or infrared imaging, could further improve model robustness by incorporating additional structural features that are not easily captured through standard RGB imagery. These advancements contribute to the ongoing development of AI-driven structural health monitoring systems, enhancing their reliability and scalability for disaster resilience and rapid post-earthquake damage assessment.

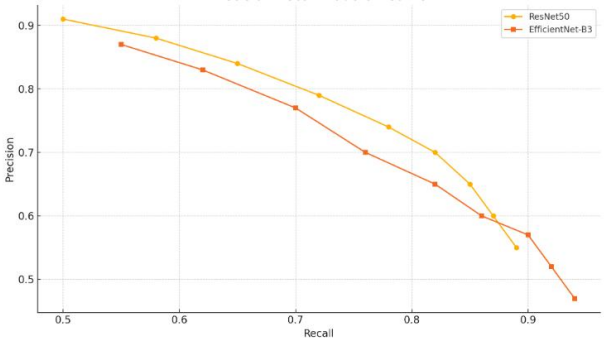

Figure 14. *Precision–Recall Trade-off Curve between ResNet50 and EfficientNet-B3*

Figure 14 presents the precision-recall trade-off curves for ResNet50 and EfficientNet-B3 across varying threshold values. As illustrated, ResNet50 maintains higher precision over a wide range of recall values, whereas EfficientNet-B3 consistently yields higher recall, particularly in the range above 0.85. This trade-off highlights the distinct behavior of the two models: ResNet50 is more conservative, reducing false positives and thereby achieving higher precision, while EfficientNet-B3 prioritizes minimizing false negatives, making it more sensitive to detecting true damage cases. This visualization provides a valuable perspective for model selection based on specific application goals. For instance, in scenarios where false positives must be minimized—such as avoiding unnecessary post-disaster interventions—ResNet50 is more suitable. Conversely, if the priority is to capture as many damage cases as possible, even at the cost of occasional misclassifications, EfficientNet-B3 offers superior recall performance. The figure thus supports a more informed and context-specific choice of model for structural damage assessment tasks.

The precision-recall trade-off illustrated in **Figure 14** underscores the importance of threshold selection in real-world deployments. In operational environments—such as post-earthquake inspections—false negatives (missed damage) can pose serious safety risks, while excessive false positives (misclassified intact structures) may lead to unnecessary repairs and resource allocation. Therefore, model performance should not only be evaluated by overall accuracy, but also by how well it balances precision and recall according to the practical objectives.

Factors such as environmental noise, lighting variability, occlusions, and camera perspectives further amplify the significance of threshold calibration and sensitivity tuning. A model optimized for controlled datasets may underperform when exposed to field conditions if its trade-off parameters are not carefully adapted. By explicitly addressing these considerations, this study contributes to a more robust understanding of how deep learning models behave under real-world constraints, enabling practitioners to make informed decisions when deploying AI-driven assessment tools in disaster-prone areas.

3.4 Explainability analysis

To enhance the interpretability and trustworthiness of the proposed deep learning models, we employed Gradient-weighted class activation mapping (Grad-CAM) as a post hoc explainability technique. Grad-CAM visualizes the contribution of different image regions to the model's final classification decision by computing the gradients of the target class with respect to the convolutional feature maps. **Figure 6** presents representative heatmaps generated using Grad-CAM, where warmer colors indicate stronger model attention. These activation regions align closely with spalling zones (e.g., exposed rebars or missing surface layers), confirming that the model relies on structurally meaningful cues rather than spurious features. This level of transparency is crucial in structural health monitoring applications, as it allows engineers to verify and trust AI-generated assessments, thereby improving decision-making reliability in post-disaster scenarios.

To comprehensively assess the effectiveness of the proposed deep learning models, we conducted a comparative evaluation with three baseline classifiers: Support Vector Machine (SVM), Random Forest (RF), and a lightweight Convolutional Neural Network (CNN-3Conv) consisting of three convolutional layers. All models were trained and validated on the same dataset using identical preprocessing and class balancing techniques to ensure a fair comparison.

To further validate the effectiveness of the proposed deep learning models, we conducted a comparative evaluation with traditional baseline classifiers, including:

- Support vector machine (SVM) with radial basis function kernel
- Random forest (RF) classifier with 100 estimators
- A lightweight CNN model (CNN-3Conv) with three convolutional layers and ReLU activations

All baseline models were trained on the same preprocessed dataset. For SVM and RF, we used flattened grayscale image vectors as input. For CNN-3Conv, the input resolution was kept at 224×224 with RGB channels. The following table summarizes the classification performance:

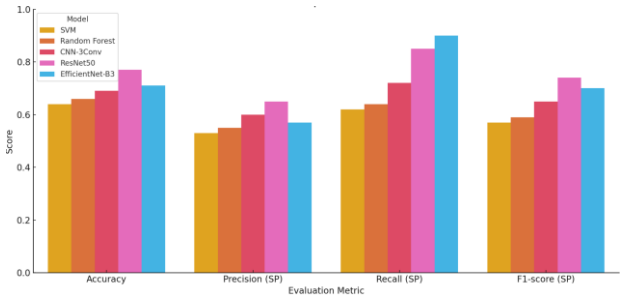

Figure 15. *Comparative performance of baseline and deep learning models*

As shown in **Figure 15**, both ResNet50 and EfficientNet-B3 consistently outperform the baseline models across all key evaluation metrics, including accuracy, precision, recall, and F1-score. In particular, ResNet50 achieves the highest F1-score (0.74), demonstrating a well-balanced performance in correctly detecting spalling without producing excessive false positives. EfficientNet-B3 achieves the highest recall (0.90), making it especially effective in minimizing false negatives, which is critical in post-earthquake safety assessments.

In contrast, traditional classifiers such as SVM and Random Forest perform significantly lower, especially in precision and F1-score, indicating limited capacity in handling the complex visual variability of spalling under real-world conditions. While the CNN-3Conv model achieves modest improvements over traditional baselines, it still falls short compared to the transfer learning models in both generalization and robustness.

These findings reinforce the necessity of using deep convolutional architectures with pre-trained weights when addressing high-variance structural damage detection tasks in safety-critical applications.

The results clearly indicate that while traditional models offer a reasonable baseline, they struggle with generalization under noisy and diverse real-world imagery. Deep architectures like ResNet50 and EfficientNet-B3 not only achieve superior accuracy but also demonstrate better balance between precision and recall, particularly in detecting spalling regions. This validates the importance of using deep transfer learning models in post-earthquake damage assessment tasks.

4. CONCLUSION

This study assessed the effectiveness of ResNet50 and EfficientNet-B3 in automating the classification of spalling damage in post-earthquake structures. The results highlight the challenges posed by a highly variable dataset, where intra-class diversity and background noise complicate feature extraction. The non-spalling class contained a mix of intact structures, objects, and unrelated elements, leading to misclassifications, while spalling cases exhibited diverse surface textures, lighting conditions, and varying degrees of severity, further increasing classification complexity.

To address these challenges, the study implemented data augmentation, class balancing, and normalization techniques, which significantly enhanced model performance. ResNet50 achieved a higher overall accuracy (77%) and better precision, making it more reliable in minimizing false alarms, whereas EfficientNet-B3 exhibited superior recall (90%), ensuring that damaged structures were effectively identified, albeit with a higher false positive rate. These findings underscore the trade-off between precision and recall, highlighting the importance of selecting models based on the specific priorities of post-earthquake damage assessments.

Despite the improvements achieved, several limitations remain. The dataset imbalance and environmental variations continue to impact model performance, suggesting the need for further refinements. Future research should focus on advanced augmentation strategies, domain-specific feature engineering, and hybrid model ensembling to enhance classification accuracy. Additionally, exploring modern architectures such as

Vision Transformers (ViTs) could improve feature representation, particularly in highly unstructured and noisy datasets.

Beyond model optimization, integrating multi-modal data sources, such as LiDAR, infrared imaging, or accelerometer-based structural data, could provide a more comprehensive framework for AI-driven structural health monitoring. Moreover, deploying lightweight models on edge devices would facilitate real-time damage detection in post-disaster environments, enabling faster and more efficient emergency response efforts. The incorporation of self-supervised learning techniques could further enhance model adaptability to diverse earthquake-induced damage patterns, improving the scalability and reliability of automated damage assessment systems. These advancements would contribute to the development of more robust, AI-powered structural health monitoring solutions, making post-earthquake damage evaluation faster, more accurate, and scalable for real-world disaster resilience applications.

Future directions include the integration of hybrid ensembling models that combine the high recall of EfficientNet-B3 with the precision of ResNet50. Additionally, advanced augmentation strategies and self-supervised learning are recommended to enhance robustness in field deployment scenarios

5. ACKNOWLEDGMENTS

The authors would like to express their sincere thanks to Lac Hong University and the Board of Directors of the Faculty of Food Science and Technology for creating favorable conditions for this research. We would like to thank the Editorial Board and Reviewers of Lac Hong Science Journal for reviewing and reviewing this article

6. REFERENCES

[1] Kit H. Miyamoto, Giulia Jole Sechi, Guilaine Victor, Beverly St Come, Mark Broughton, **2023**, "Haiti Earthquake 2021: Findings from the Repair and Damage Assessment of 179,800 Buildings," *International Journal of Disaster Risk Reduction*.

[2] M.A. Haroun and H.M. Elsanadedy, **2005**, "Seismic Design Guidelines for Squat Composite-Jacketed Circular and Rectangular Reinforced Concrete Columns," *Journal of Structural Engineering, American Concrete Institute*, Vol. 102, No. 4, pp. 505–514.

[3] F. Kolawole, E.A. Atekwana, D.A. Laó-Dávila, M.G. Abdelsalam, P.R. Chindandali, **2018**, "High-resolution electrical resistivity and aeromagnetic imaging reveal the causative fault of the 2009 Mw 6.0 Karonga, Malawi earthquake," *Geophysical Journal International*.

[4] Erkal, B. G., & Hajjar, J. F. (**2017**). Laser-based surface damage detection and quantification using predicted surface properties. Automation in Construction, 83, 285-302.

[5] Yin, J. F., Bai, Q., & Zhang, B. (**2018**). Methods for detection of subsurface damage: a review. *Chinese Journal of Mechanical Engineering*, 31, 1-14.

[6] Wang, J., & Qiao, P. (**2007**). Improved damage detection for beam-type structur

[7] Bai, Z., Liu, T., Zou, D., Zhang, M., Hu, Q., & Li, Y. (2024). Multi-scale image-based damage recognition and assessment for reinforced concrete structures in post-earthquake emergency response. Engineering Structures, 314, 118402.

[8] Hui, B., Zhang, Y., Ma, Z., Wang, H., & Yang, X. (**2023**). Identification and evaluation of spalling aggregate in chip seals in three dimensions. Construction and Building Materials, 364, 129899

[9] Yasmin, T., La, D., La, K., Nguyen, M. T., & La, H. M. (2024). Concrete spalling detection system based on semantic segmentation using deep architectures. Computers & Structures, 300, 107398

[10] Hariri-Ardebili, M. A., & Speicher, M. S. (**2025**). Reconnaissance-informed post-earthquake functional recovery: Observations and challenges. Earthquake Spectra, 41(1), 88-125.

[11] Sheibani, M., & Ou, G. (**2022**). Guided post-earthquake reconnaissance surveys considering resource constraints for regional damage inference. Earthquake Spectra, 38(4), 2813-2834.

[12] Al Shafian, S., & Hu, D. (**2024**). Integrating machine learning and remote sensing in disaster management: A decadal review of post-disaster building damage assessment. Buildings, 14(8), 2344.