ELECTRICAL ENGINEERING AND ENERGY

OPEN ACCESS

Performance Evaluation of Deep Learning Architectures for Tile Defect Detection

Ayça Demir, Humar Kahramanlı Örnek

Department of Computer Engineering, Selçuk University, Konya, Türkiye

Submitted: 16 November 2025
Accepted : 13 December 2025
Online First: 16 December 2025

Corresponding author
Humar Kahramanlı Örnek,
hkahramanli@selcuk.edu.tr

DOI:10.64470/elene.2025.1016

Abstract: In this study, an artificial intelligence-based quality control system was developed for the automatic detection and classification of defects in a tile. The dataset created to reduce human-induced errors in the production process and increase inspection accuracy consists of a total of 405 images.

During the model development phase, CNN, MobileNetV2, ResNet50, and EfficientNetB0 architectures were used. The performance of the models was evaluated using the 10-fold cross-validation method for an objective comparison.

The experimental results show that the EfficientNetB0 architecture achieved the highest performance with an accuracy rate of 96.73%. ResNet50 achieved 95.45%, CNN achieved 94.91%, and MobileNetV2 achieved 92.36% accuracy.

Keywords Deep Learning, Image Processing, Tile Defect Detection, Transfer Learning.

1. Introduction

Ceramic tiles have adorned buildings for centuries as a decorative element in both the interior and exterior of various architectural structures, including residential buildings, public works, and places of worship. Their history dates back to ancient civilizations such as Egypt, Assyria, and Babylonia, and later, the Romans and Greeks also used decorative tiles. The spread of Islam further accelerated their cultural significance, and eventually, they became widespread in Europe during the Middle Ages (Michalak, 2021). The Industrial Revolution made tiles widely available. Ceramic tiles have been valued for centuries due to their durability and excellent resistance to external elements (Michalak, 2021). Today, modern ceramic tile manufacturing has reached remarkable levels of mechanization and automation, particularly in raw material processing (Lu et al., 2022).

The quality of ceramic tiles is one of the key factors determining the competitiveness of building materials companies. Even minor defects on the surface or edges of tiles can reduce their aesthetic and performance properties, as well as lead to significant economic losses. The production process of ceramic tiles, on the other hand, can lead to various defects, such as dirt, scratches, holes, uneven colors, corner errors, and so on. However, despite technological advances, the inspection of finished tiles for defects is largely done manually. Traditional visual inspection methods, performed manually by operators, are characterized by low speed, high subjectivity, and staff fatigue. This traditional approach is not only labor-intensive but also prone to human error, highlighting the need for more efficient and intelligent quality control solutions (Lu et al., 2022). In this regard, automated visual inspection systems (AVIS) based on machine vision and artificial intelligence technologies have been actively developed in recent years.

Classic image processing algorithms using threshold binarization, morphological operations, and edge detection methods made it possible to identify surface defects such as cracks, stains, glaze bubbles, and scratches. However, they demonstrated low resistance to changes in lighting, texture, and tile geometry. In addition, most research has focused on surface defects, while edge (boundary) defects—curvature, chipping, and thickness deviations—remained understudied, despite their critical impact on installation quality and product durability. The development of deep learning has opened up new possibilities for automatic image analysis. Modern architectures, such as Convolutional Neural Networks (CNN) and YOLO, have demonstrated high accuracy and generalization ability in defect detection and classification tasks. However, the application of deep models for ceramic tile edge defect detection tasks remains rare. This work aims to develop a hybrid system that combines the advantages of deep learning and classical geometric analysis methods for accurate, fast, and cost-effective real-time detection of defects on the edges and surface of ceramic tiles. Research in the field of automatic quality control of ceramic tiles can be divided into three stages:

- The use of traditional image processing methods.
- The development of geometric analysis algorithms for measuring dimensions and distortions.
- The introduction of deep learning methods and hybrid intelligent systems.

In the first stage, Elbehiery et al. (2005) proposed a method for detecting surface defects based on morphological operations and the Canny operator. Their approach improved the detection of defect boundaries but proved to be sensitive to noise and unstable when lighting conditions changed. Later, Hocenski and Keser (2007) applied an improved Canny algorithm and a derivative of directional gradients to detect edge defects, achieving an accuracy of about 98% in laboratory tests. Golkar et al. (2011) paid particular attention to edge defects, proposing the AVIS system, which uses inexpensive CMOS cameras and LED lighting to detect warping, chipping, and thickness deviations in tiles. Their approach was based on extracting linear features and comparing them with reference lines, which allowed them to achieve a relative measurement error of about 1.44%. However, the method depended on the fixed position of the cameras and required precise threshold settings, which limited its application on production lines.

With the advent of deep convolutional neural networks, significant progress has been made in automatic image analysis. ResNet, EfficientNet, U-Net, and YOLOv8 architectures provide high accuracy in defect classification and segmentation even in noisy and unstructured data conditions. However, most current research focuses on surface defects, while the analysis of geometric deviations, such as edge bending or thickness heterogeneity, remains understudied.

In recent years, a number of authors have proposed hybrid approaches that combine geometric approximation of lines with neural network training. Such systems use CNN to locate potential defect areas, then perform quantitative analysis of shape and curvature using algorithms such as Hough and PCA. This class of methods demonstrates increased accuracy and noise resistance, combining the interpretability of traditional image processing with the adaptability of deep models.

Thus, analysis of the literature shows that the development of a universal, stable, and cost-effective system for the automatic detection of defects in ceramic tile edges, integrating the advantages of classical and neural network analysis methods, remains a pressing scientific task.

In the second chapter a framework of this study and related background theory are presented. The results of the experiments and evaluation of these results are presented in the third chapter. In the final chapter this paper is concluded.

2. Material and Method

The dataset used in this study was created from images of defective and intact tile surfaces obtained on the production line. The images were collected to represent various lighting conditions, color tones, and surface textures. All images were captured using a camera mounted on a fixed rod, with the tiles placed on a black surface at a fixed height and acquired sequentially under controlled settings. The dataset contains a total of 405 images, including 300 defective and 105 intact tile images, resulting in an initially imbalanced class distribution. To mitigate the potential impact of this class imbalance on model performance and to improve generalization, data augmentation techniques were applied more intensively to the minority class. Augmentation operations such as random rotation, horizontal and vertical flipping, brightness adjustment, noise injection, and zooming were performed. Through this targeted augmentation strategy, the dataset was expanded to 550 images, comprising 300 defective and 250 intact samples, thereby significantly reducing the degree of class imbalance. In addition, class-aware evaluation was conducted using metrics beyond overall accuracy to ensure that the reported performance reflects balanced behavior across both classes. Figure 1 presents representative examples of defective and intact tiles.

Figure 1 The samples of defective and intact tiles

In this study, four different deep learning models: CNN, MobileNetV2, ResNet50, EfficientNetB0 were used for the classification of tile defects.

2.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are artificial neural network architectures based on local feature extraction, widely used in the field of deep learning, particularly for image processing, object recognition, and classification tasks. CNNs scan the input data using sliding filters (kernels) to detect important patterns in local regions, thereby creating feature maps. This approach allows the network to automatically learn patterns itself, replacing the manually defined features used in traditional methods (Li et al., 2020). CNN architectures typically consist of convolutional, activation, pooling, and fully connected layers, and these structures enhance the model's overall performance by enabling hierarchical feature extraction (Khan et al., 2019). As noted by LeCun et al. (2015), deep convolutional networks have achieved significant success on complex data types such as images, video, and audio. Recent studies, however, have examined in detail the effects of developments in convolution types (e.g., 1D, 2D, or multi-dimensional), arrangement techniques, and architectural designs on CNN performance (Li et al., 2020; Younesi et al., 2024).

2.2 MobileNetV2

MobileNetV2 is an efficiency-focused CNN architecture designed for use on mobile and embedded devices with limited resources. The key innovation of this architecture is that it is constructed using inverted residual expansion-projection layers, as opposed to traditional residual blocks; processing is performed with depthwise convolutions while keeping the input and output channels narrow, and it is recommended to remove non-linearities in the bottleneck layers (Sandler et al., 2018). This design enables processing with fewer multiply-adds operations and a lower number of parameters while maintaining the same level of accuracy; for example, it is more efficient in object classification, detection, and segmentation tasks (Sandler et al., 2018).

2.3 ResNet50

ResNet50, a significant milestone in the field of deep learning, is a CNN architecture based on the "residual learning" approach. Developed by Kaiming He and colleagues, this model provides a solution to the vanishing gradient and accuracy drop problems encountered in multi-layer networks (He, et al., 2016). The fundamental difference of ResNet50 is that it uses "shortcut connections" instead of traditional layer stacks, allowing the outputs of certain layers to be directly transferred to subsequent layers. This enables the network to maintain its learning performance while increasing its depth, making the training process more stable. With its 50-layer structure, the model has achieved high accuracy rates on large datasets such as ImageNet and has served as the basis for many deep network architectures developed later.

2.4 EfficientNetB0

EfficientNetB0 is an innovative convolutional neural network architecture that aims to optimize the balance between model accuracy and computational efficiency in the field of deep learning. Proposed by Tan and Le (2019), this approach is based on the "compound scaling" method, which simultaneously scales model width, depth, and resolution. This method allows for the development of models that are lightweight enough to run on smaller devices yet powerful enough to deliver high accuracy by balancing resource usage. EfficientNetB0 has achieved higher accuracy with fewer parameters compared to other CNN-based architectures and has thus begun to be widely used in many areas such as classification, object recognition, and medical image analysis (Tan & Le, 2019).

In this study, the confusion matrix and a set of evaluation metrics to understand how well the classification model performed were used. For each class, a matrix was formed by placing the actual observations and the model's predictions side by side, as shown in Table 1. In this structure, each column reflects the predicted class, while each row reflects the actual class. A true positive (TP) and a true negative (TN) mark the moments when the model recognizes the tiles as they truly are: identifying a condition when it is defective or recognizing its intact. False positives (FP) and false negatives (FN), on the other hand, remind us that scientific tools are not perfect. A false positive occurs when the model signals danger where there is none, while a false negative hides a condition that truly exists. Together, the quantities TP, FP, TN, and FN allow us to compute essential indicators such as accuracy, sensitivity, precision, the Area Under the Curve (AUC), and the F1-Score. These measures help us judge the reliability and usefulness of the model's conclusions.

Table 1 Confusion matrix definition

Reference	Predictions	
	Defective	Intact
Defective	TP	FN
Intact	FP	TN

To evaluate the model's performance with rigor, the 10-fold cross-validation method was employed. In this approach, the dataset was gently separated into ten equal portions. During each cycle, nine of these portions served as the training ground where the model learned, while the remaining part was reserved for testing its understanding. This sequence was repeated ten times, allowing the model to be tested under many different conditions. Each fold offered its own accuracy value and its own confusion matrix, much like repeated experiments that reveal subtle variations in a phenomenon. Figure 2 illustrates this process, showing how the rotation of training and testing parts helps us reach a more trustworthy evaluation of the model's true capability (Abdullayeva & Kahramanlı Örnek, 2024).

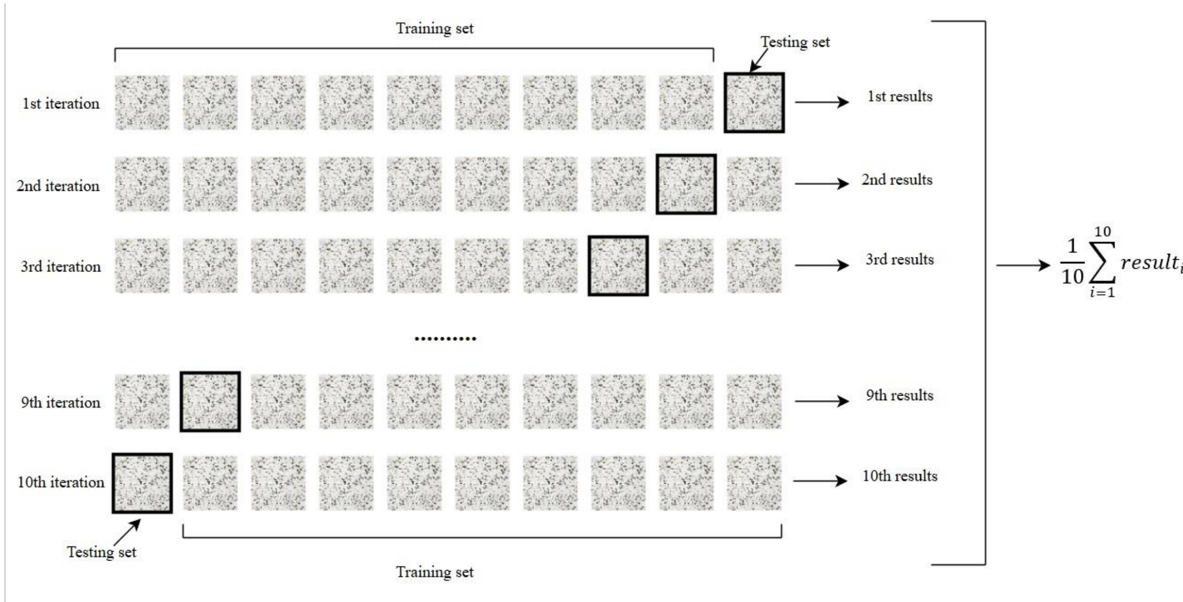


Figure 2 Cross validation process

A confusion matrix was employed to observe how well the model performed for each class, using measures such as Accuracy, Precision, Sensitivity, F1-Score, and AUC. These metrics help us look closely at the strengths and weaknesses of the model, much like examining the results of repeated experiments. The formulas for each of these measurements are presented in Table 2, offering a clear view of how they are calculated and interpreted.

Table 2 Classification evaluation metrics

Metric	Formula
Accuracy	$Acc = \frac{tp + tn}{tp + fp + tn + fn}$
Precision	$P = \frac{tp}{tp + fp}$
Sensitivity	$S = \frac{tp}{tp + fn}$
F1-Score	$F1 - score = \frac{2\left(\frac{tp}{tp + fn}\right)\left(\frac{tp}{tp + fp}\right)}{\left(\frac{tp}{tp + fn}\right) + \left(\frac{tp}{tp + fp}\right)}$
AUC	$AUC = \frac{1}{2}\left(\frac{tp}{tp + fn} + \frac{tp}{tp + fp}\right)$

*tp: true positive, tn: true negative, fp: false positive, fn: false negative

3. Results and Discussion

This section presents experimental results to evaluate the performance of the developed system. Four different deep learning models—CNN, MobileNetV2, EfficientNetB0, and ResNet50—were trained and tested using a 10-fold cross-validation method. The confusion matrix values generated for each fold allowed for detailed analysis of the models' correct classification rates and error types. All models used in the study were trained using the transfer learning approach. This reduced the training time and increased the accuracy rate. In Table 3, the accuracy, precision, sensitivity, F1-score, and AUC values of all used algorithms are given. Figure 3 shows the confusion matrix of all four methods.

The results of the 10-fold cross-validation applied to measure the generalization ability of the model are presented in the bar chart in Figure 4.

The classification accuracy of the CNN is 94.91%. The lowest accuracy is 89.09% in the third and sixth folds, while it reaches 100% accuracy in the eighth fold. The standard deviation of the fold accuracies is 3.72.

The classification accuracy of the MobileNetV2 is 92.36%. The lowest accuracy is 78.18% in the seventh fold, while it reaches 100% accuracy in the second and fourth folds. The standard deviation of the fold accuracies is 7.80.

The classification accuracy of the ResNet50 is 95.45%. The lowest accuracy is 90.91% in the second and sixth folds, while it reaches 98.18% accuracy in five folds. The standard deviation of the fold accuracies is 3.12.

The classification accuracy of EfficientNetB0 is 96.73%. The lowest accuracy is 90.91% in the last fold, and it achieves 94.55% accuracy in the first fold. In three folds, the method achieves 96.36% accuracy, while in five folds it reaches 98.18%. The standard deviation of the fold accuracies is 2.54.

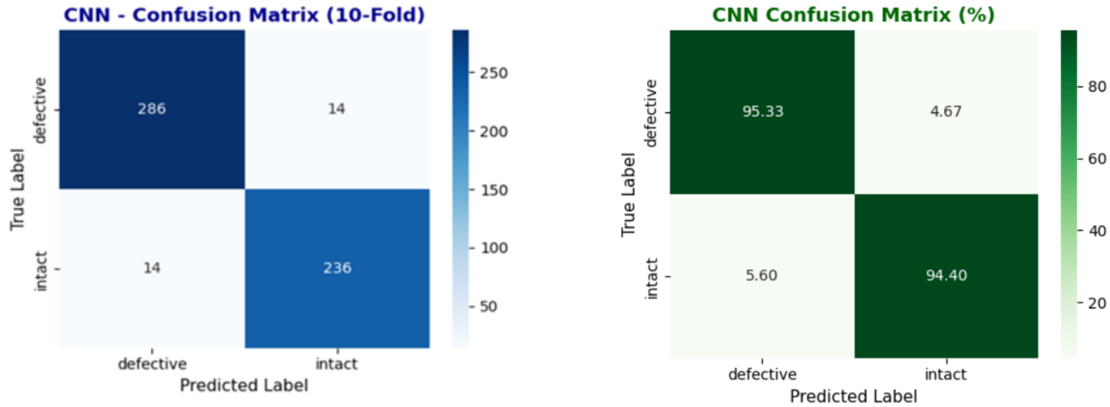
Sensitivity, Precision, F1-score, and AUC are all 0.95, 0.96, and 0.97 for CNN, ResNet50, and EfficientNetB0, respectively. It demonstrates the stability of these algorithms for this problem.

As seen in Table 3, Figure 3, and Figure 4, all four algorithms achieved satisfactory results, with EfficientNetB0 obtaining the highest classification accuracy at 96.73%. Figure 4 also shows that EfficientNetB0's performance across folds is more balanced compared to the other models.

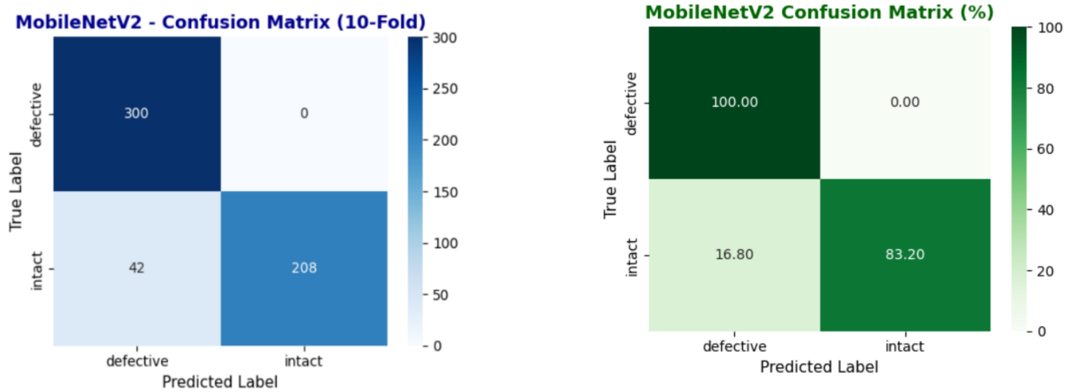
MobileNetV2 performed the worst among the models. This algorithm misclassified more tiles as defective when they were not, resulting in a higher number of false positives. This resulted in high sensitivity as 1 and low precision as 0.88 values (Table 3). Figure 4 shows that MobileNetV2's performance across folds is also more imbalanced.

Table 3 Accuracy, precision, sensitivity, F1-score, and AUC values of all algorithms

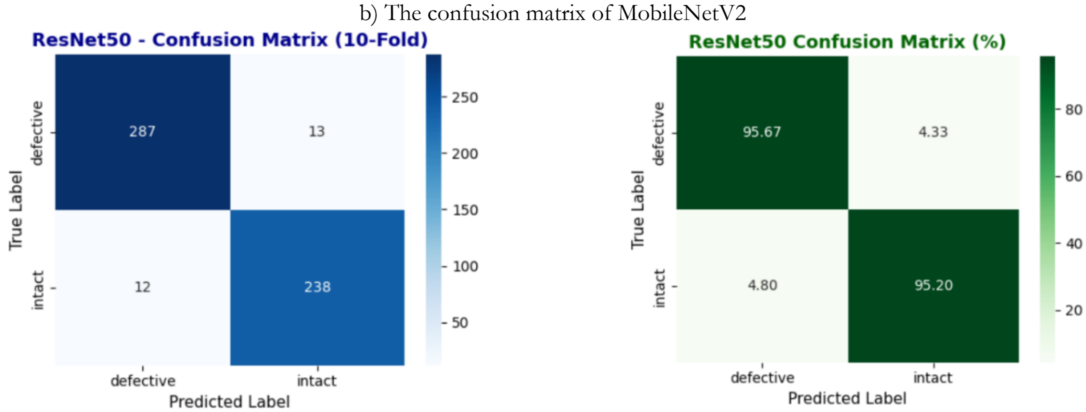
	Accuracy	Precision	Sensitivity	F1-score	AUC
CNN	94.91	0.95	0.95	0.95	0.95
MobileNetV2	92.36	0.88	1	0.94	0.94
ResNet50	95.45	0.96	0.96	0.96	0.96
EfficientNetB0	96.73	0.97	0.97	0.97	0.97



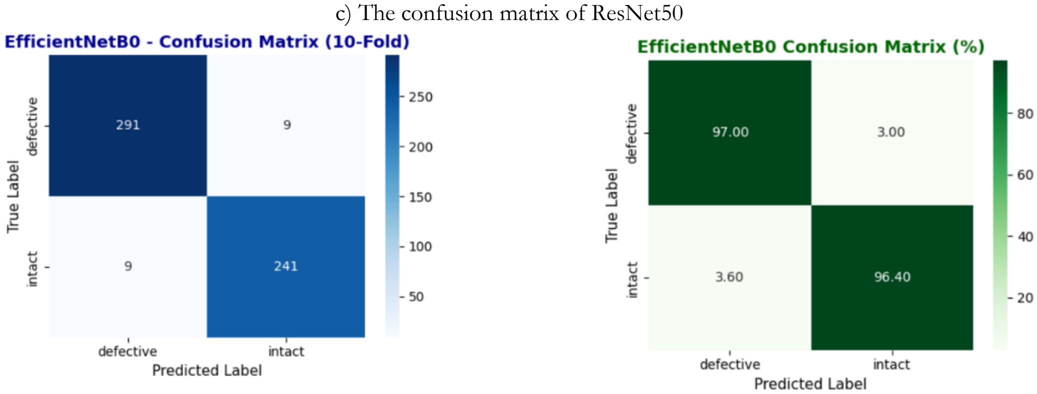
a) The confusion matrix of CNN



b) The confusion matrix of MobileNetV2

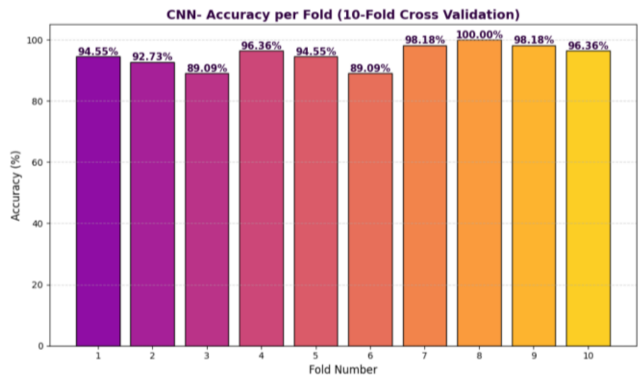


c) The confusion matrix of ResNet50

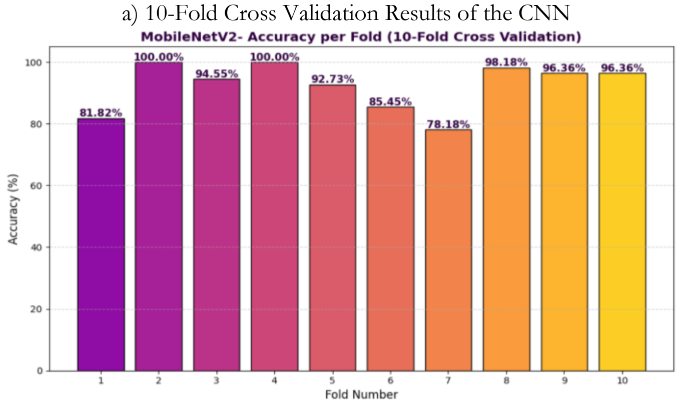


d) The confusion matrix of EfficientNetB0

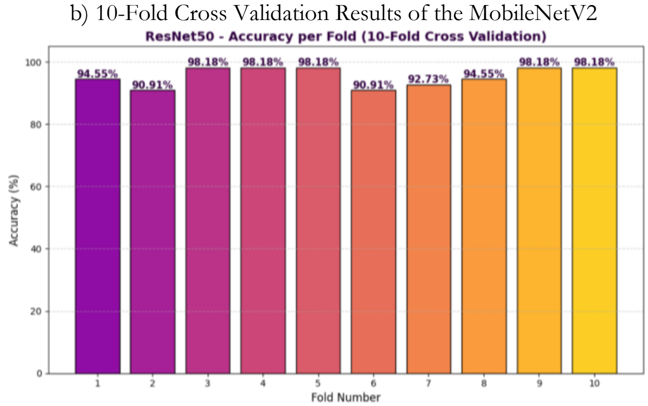
Figure 3. Confusion matrix of methods



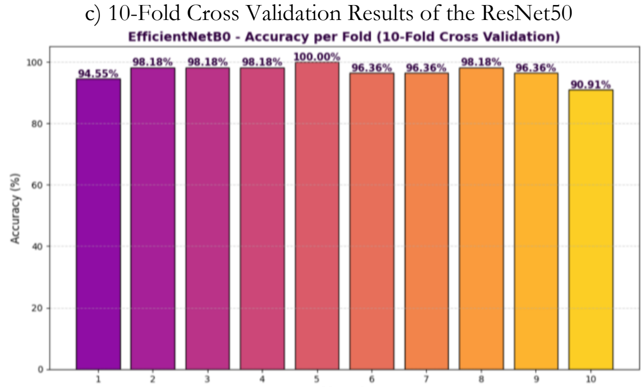
a) 10-Fold Cross Validation Results of the CNN



b) 10-Fold Cross Validation Results of the MobileNetV2



c) 10-Fold Cross Validation Results of the ResNet50



d) 10-Fold Cross Validation Results of the EfficientNetB0

Figure 4 10-Fold Cross Validation Results of all models

4. Conclusion

In this study, the quality of ceramic tiles was analyzed. Four deep learning architectures—CNN, MobileNetV2, ResNet50, and EfficientNetB0—were evaluated using k-fold cross-validation. The results demonstrate that all models achieved satisfactory performance; however, EfficientNetB0 outperformed the others with the highest accuracy and the lowest performance variability across folds. EfficientNetB0 showed strong generalization capability, making it the most suitable architecture for tile defect detection within the scope of this study.

Overall, the findings highlight the potential of deep learning-based approaches for automated tile quality inspection. Future work may include expanding the dataset, experimenting with additional architectures, and integrating real-time inference to further support industrial deployment.

Declaration of Ethical Standards

As the authors of this study, we declare that he complies with all ethical standards.

Credit Authorship Contribution Statement

A.Demir: Software, Methodology, Validation, Formal analysis, Writing -Original Draft, Visualization.
H.Kahramanlı Örnek: Investigation, Writing, Review & Editing, Supervision.

Declaration of Competing Interest

The authors declared that they have no conflict of interest.

Funding / Acknowledgements

The authors declare that this study has not received any funding.

Data Availability

The dataset used in this study originates from Ayça Demir's senior capstone project. After the article is published, the data will be made available for researchers.

References

Abdullayeva, E. & Kahramanli Örnek, H. (2024). Diagnosing Epilepsy from EEG Using Machine Learning and Welch Spectral Analysis, *Traitement du Signal* Vol. 41, No. 2, April, 2024, pp. 971-977, https://doi.org/10.18280/ts.410237

Elbehiery, H., Hefnawy, A. and Elewa, M. (2007). Surface Defects Detection for Ceramic Tiles Using Image Processing and Morphological Techniques, *World Academy of Science, Engineering and Technology*, 5

Golkar, E., Patel, A., Yazdi, L. & Prabuwono, A.S. (2011). Ceramic Tile Border Defect Detection Algorithms in Automated Visual Inspection System, *Journal of American Science*, 7(6).

He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778, https://doi.org/10.1109/CVPR.2016.90.

Hocenski, Z & Keser, T. (2007). Failure detection and isolation in ceramic tile edges based on contour descriptor analysis, *Mediterranean Conference on Control & Automation*, Athens, Greece, pp. 1-6, https://doi.org/10.1109/MED.2007.4433713.

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2019). A survey of the recent architectures of deep convolutional neural networks. arXiv preprint arXiv:1901.06032.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning. Nature*, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Li, Y., Yang, T., Peng, S., & Liu, Y. (2020). A survey of convolutional neural networks: Analysis, applications, and prospects. arXiv preprint arXiv:2004.02806.

Lu, Q., Lin, J., Luo, L., Zhang, Y. & Zhu, W. (2022). A supervised approach for automated surface defect detection in ceramic tile quality control, *Advanced Engineering Informatics*, 53, 101692. https://doi.org/10.1016/j.aei.2022.101692

Michalak, J. (2021). Ceramic Tile Adhesives from the Producer's Perspective: A Literature Review. *Ceramics*, 4, 378–390. https://doi.org/10.3390/ceramics4030027

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. arXiv preprint arXiv:1801.04381

Tan, M. & Le, Q.V. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proceedings of the 36th International Conference on Machine Learning*, 6105-6114.

Younesi, E., Ansari, R., Fazli, M., Ejlali, A., Shafique, M., & Henkel, J. (2024). A comprehensive survey of convolutions in deep learning. arXiv preprint arXiv:2402.15490.