Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Prediction of Aggregate Particle Roundness Classes Using Deep Learning and Comparison of CNN Models

Gokhan Ucar

Tatvan Vocational School, Bitlis Eren University

Fatih Isık

fatih.isik@atauni.edu.tr

Atatürk University

Article

Keywords: Aggregate morphology, image processing, deep learning, convolutional neural network, particle shape

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11649358/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Prediction of Aggregate Particle Roundness Classes Using Deep Learning and Comparison of CNN Models

Gokhan Ucar [1] **and Fatih Isık** [2]

Abstract

The morphological characteristics of aggregate particles, including shape, angularity, and roundness, play a crucial role in determining interparticle contact, mechanical interlocking, and frictional behavior, thereby influencing the engineering performance of granular materials. This study investigated the classification of aggregate morphological groups using deep learning and image-based analysis. Crushed aggregates were subjected to Los Angeles abrasion tests at five cycle levels (0, 500, 1000, 2250, and 6000). Based on preliminary morphological assessments, the 0-, 1000-, and 6000-cycle groups were selected to represent three more distinguishable aggregate morphological categories: Angular, Subrounded, and Well Rounded, respectively. A dataset of 4500 original particle images was expanded to 9000 images through 90° rotation-based data augmentation. Six convolutional neural network architectures: DenseNet-201, MobileNet-V2, ResNet-18, ResNet-50, ResNet-101, and SqueezeNet were evaluated using an 80:20 training–validation split. Model generalization was assessed using an independent test set comprising 300 images, with 100 images per class. DenseNet-201 achieved the highest independent test accuracy of 91.33%, compared with its validation accuracy of 99.28%. The differences between validation and independent test performance emphasize the importance of independent evaluation when assessing CNN generalization. These findings demonstrate the potential of deep learning for distinguishing experimentally defined, abrasion-associated aggregate morphological groups.

Keywords: Aggregate morphology, image processing, deep learning, convolutional neural network, particle shape.

[1] Construction Department, Tatvan Vocational School, Bitlis Eren University, Bitlis, Turkey

[2] Construction Technology, Vocational School of Technical Sciences, Atatürk University, Erzurum, Turkey

Corresponding author(s)

Fatih Isık, fatih.isik@atauni.edu.tr

1. Introduction

Aggregate morphology, including particle form, angularity, roundness, and surface texture, influences interparticle contact, mechanical interlocking, and frictional behavior[1,2]. Variations in roundness and angularity consequently affect the mechanical behavior of granular materials[3], while these morphological characteristics represent distinct components of aggregate shape[4]. Therefore, objective, quantitative, and repeatable methods are essential for aggregate morphological characterization[2,5,6].

Image-based techniques enable rapid and objective quantification of aggregate form, angularity, surface texture, and dimensional characteristics[5,6,7]. Automated imaging systems and three-dimensional approaches further support the characterization of fine and coarse aggregates[4,8]. Recent advances in computer vision and deep learning have extended image-based methods to automated particle identification and morphological analysis[9-11]. In particular, convolutional neural networks (CNNs) can learn discriminative image features for classification and characterization tasks[12,13]. Related applications include image-based gravel particle characterization[14] and CNN-based classification of volcanic ash particle shapes[15].

Convolutional neural networks (CNNs) are widely used in image classification because they automatically learn hierarchical and discriminative features from visual data[16]. Different CNN architectures vary in depth, connectivity, and computational cost, with ResNet employing residual connections to facilitate deep network training[17]. Deep learning approaches have also demonstrated potential for characterizing granular particle shapes and predicting geometric properties such as roundness[18]. However, limited training data may lead to overfitting and poor generalization, which can be mitigated through data augmentation techniques such as rotation, translation, cropping, and reflection[19]. Nevertheless, studies comparing CNN architectures for classifying angular, subrounded, and well-rounded aggregates using independent test datasets remain limited. Furthermore, deep learning-based classification of abrasion-induced changes in aggregate roundness following Los Angeles abrasion has received insufficient attention.

The objective of this study was to classify aggregate particles exhibiting distinct degrees of roundness following the Los Angeles abrasion process using deep learning-based image analysis and to compare the predictive performance of different CNN architectures. Crushed aggregates were subjected to five abrasion cycle

levels (0, 500, 1000, 2250, and 6000). Based on preliminary morphological assessments, the 0-, 1000-, and 6000-cycle groups were selected to represent three distinguishable roundness categories: Angular, Subrounded, and Well Rounded, respectively. A total of 4500 original aggregate images were acquired, comprising 1500 images per class. Through 90° rotation-based data augmentation, the dataset was expanded to 9000 images (3000 per class). To prevent data leakage, each original image and its rotated counterpart were assigned to the same data subset before an 80:20 training–validation split. Six CNN architectures—DenseNet-201, MobileNet-V2, ResNet-18, ResNet-50, ResNet-101, and SqueezeNet—were evaluated. Model generalization was assessed using an independent test set comprising 300 images (100 distinct particles per class), entirely separate from the training and validation datasets. Differences in classification outcomes were examined through 15 pairwise McNemar tests[20], with Holm correction applied to account for multiple comparisons[21].

2. Materials and methods

2.1. Aggregate sample and Los Angeles abrasion test

The crushed aggregates were obtained from Adabağ A.Ş. (Tatvan, Bitlis, Türkiye). Los Angeles abrasion tests were conducted in accordance with ASTM C131[22] to investigate abrasion-associated changes in particle morphology. Independent 7 kg aggregate samples were subjected to five abrasion levels (0, 500, 1000, 2250, and 6000 cycles), with the 0-cycle group representing the unabraded condition. Following abrasion, aggregate particles from each group were prepared for preliminary image-based morphological assessment. Based on these assessments, the 0-, 1000-, and 6000-cycle groups were selected for subsequent CNN classification as Angular, Subrounded, and Well Rounded, respectively.

2.2. Imaging setup and image acquisition

Aggregate particles were photographed using a fixed imaging setup comprising a Nikon D500 camera equipped with a 105 mm lens. Camera height, shooting angle, and lighting conditions were kept constant throughout image acquisition to ensure consistent imaging conditions. Initially, 100 particles from each of the five abrasion levels (0, 500, 1000, 2250, and 6000 cycles) were imaged and analyzed using a MATLAB-based image processing algorithm. Particle roundness was calculated according to the method proposed by Al-Rousan[23]. Based on the preliminary morphological assessment, three abrasion groups were selected and assigned group-level morphological labels: Angular (0 cycles), Subrounded (1000 cycles), and Well Rounded (6000 cycles). Subsequently, 1500 distinct particles from each selected abrasion group were individually photographed under identical imaging conditions, producing 4500 original images (1500 per class). Class labels were assigned according to the corresponding abrasion group rather than individual roundness measurements of all 4500 particles.

2.3. Determination of roundness classes using the Al-Rousan approach

Particle roundness was determined using the image-based approach proposed by Al-Rousan[23]. Aggregate images were processed using MATLAB R2026a (The MathWorks, Natick, MA, USA) to separate particles from the background and calculate their projected area (A) and perimeter (P). Roundness was then calculated using Equation (1):

$$R = \frac{P^2}{4.\pi.A} \quad (1)$$

Where R is the roundness index, P is the particle perimeter, and A is the projected area. A value of 1 represents a perfect circle, whereas higher values indicate increasing shape irregularity and angularity[23]. Initially, 100 particles from each of the five Los Angeles abrasion levels were analyzed, and their roundness values were evaluated with reference to the classification ranges proposed by Al-Rousan[23], considering the relationship between particle shape characteristics and shape indices investigated by Arasan et al.[24]. Based on this preliminary assessment, three abrasion-associated morphological categories were defined for the study: Angular, Subrounded, and Well Rounded.

Accordingly, the particles subjected to 0, 1000, and 6000 abrasion cycles were assigned to the Angular, Subrounded, and Well Rounded classes, respectively. These abrasion-based morphological categories were subsequently used as class labels for the image dataset. An example of the MATLAB-based image processing stages used for particle identification and roundness analysis is presented in Fig. **1.**

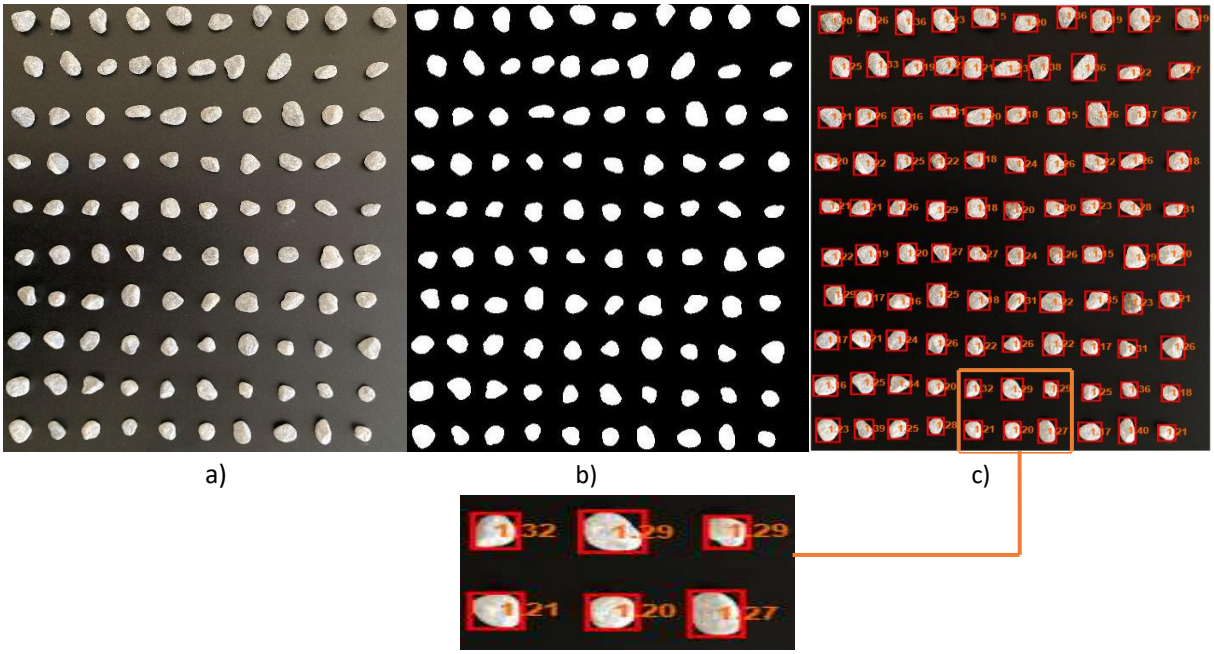


Figure 1. Image processing stages of aggregate particles according to the Al-Rousan approach: (a) raw image, (b) segmentation result, and (c) aggregate particles belonging to the well rounded class.

2.4. Dataset construction and data augmentation

Following roundness classification using the Al-Rousan approach, 1500 distinct aggregate particles per class were individually photographed, yielding 4500 original images. Data augmentation was performed in MATLAB by generating a 90° rotated version of each image, increasing the dataset to 9000 images (3000 per class). To prevent data leakage, the training–validation split was performed at the original-image level. For each class, 1200 original images were assigned to training and 300 to validation, with their rotated counterparts retained in the same subsets. Consequently, the training and validation datasets contained 7200 and 1800 images, respectively. An independent, non-augmented test set comprising 300 original images (100 distinct particles per class) was used to evaluate the generalization performance of the six CNN models. The overall experimental and computational workflow is illustrated in Fig. **2**.

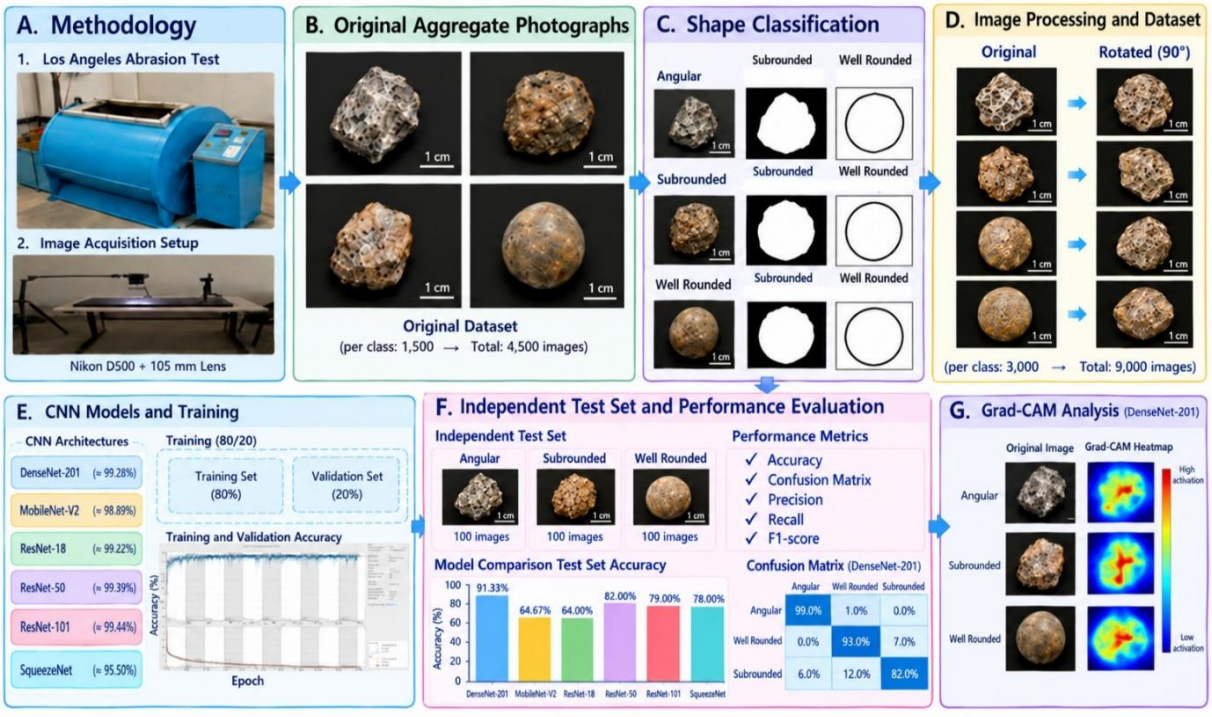


Figure 2. Experimental and Deep Learning-Based Workflow for Aggregate Roundness Classification

2.5. Deep learning models and training process

To evaluate the performance of convolutional neural networks (CNNs) in classifying aggregate particle roundness, six architectures with distinct structural characteristics and computational complexities were comparatively investigated: DenseNet-201, MobileNet-V2, ResNet-18, ResNet-50, ResNet-101, and SqueezeNet. DenseNet-201 employs dense interlayer connections that facilitate feature reuse across successive layers. MobileNet-V2 incorporates depthwise separable convolutions and inverted residual blocks to reduce computational demands while maintaining a lightweight architecture. ResNet-18, ResNet-50, and ResNet-101 represent architectures of increasing depth that employ residual connections to facilitate network optimization. SqueezeNet was included as a parameter-efficient architecture designed to achieve effective classification performance with a reduced model size. The final classification layers of all six architectures were adapted to distinguish among three aggregate roundness classes: angular, subrounded, and well rounded. Input images were resized to meet the dimensional requirements of each network. All models were trained in MATLAB using identical training and validation datasets to ensure a consistent comparative framework. Training was performed using stochastic gradient descent with momentum (SGDM) for 10 epochs, with a mini-batch size of 32. Model performance during training was monitored using the validation dataset. Following training, the generalization performance of each architecture was assessed using an independent test dataset comprising 300 original images that were excluded from both training and validation. This evaluation enabled a direct comparison of the six CNN architectures in classifying previously unseen aggregate particles according to their roundness characteristics.

2.6. Statistical analysis

The McNemar test was used to assess statistically significant differences in the classification outcomes of the six CNN models evaluated on the same independent test dataset comprising 300 images[20]. This paired test compares the correct and incorrect predictions of two classifiers on identical test instances. For each model pair, *b* represented the number of images correctly classified by the first model but incorrectly classified by the second, whereas *c* represented the reverse outcome. Images classified either correctly or incorrectly by both models were excluded from the discordant pairs. Statistical significance was assessed based on the difference between *b* and *c*. A total of 15 pairwise comparisons [6(6−1) /2] were performed, with two-sided *p*-values calculated for each comparison. To control the family-wise error rate associated with multiple testing, the Holm correction was applied[21]. Differences were considered statistically significant when the Holm-adjusted *p*-value was below 0.05. The McNemar test was used to complement, rather than replace, conventional classification metrics, including Accuracy, Precision, Recall, and F1-score.

2.7. Independent test dataset and performance evaluation

The generalization performance of the six CNN architectures was evaluated using an independent test dataset comprising 300 original images (100 per roundness class) from aggregate particles excluded from training and validation. No data augmentation was applied to the test images. Overall classification performance was assessed using accuracy, while confusion matrices were constructed to examine class-specific predictions and misclassification patterns. These analyses enabled a comparative assessment of the models' classification performance across the angular, subrounded, and well rounded classes.

2.8. Grad-CAM analysis

Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to interpret the classification decisions of DenseNet-201, the best-performing model, by visualizing image regions contributing to its predictions[25]. Model interpretability is particularly relevant in engineering applications, where understanding the features underlying predictions complements conventional performance evaluation[26]. Grad-CAM analysis was conducted on correctly classified images from the independent test dataset. Class-specific activation maps were generated and overlaid on the original aggregate images, with cool and warm colors indicating lower and higher activation levels, respectively. These visualizations provided qualitative insight into the image regions and morphological features associated with the model's classification decisions.

3. Results and discussion

3.1. Effect of Los Angeles abrasion cycles on the roundness characteristics of aggregates

To evaluate abrasion-induced changes in aggregate morphology, images of 100 particles from each of the 0, 1000, and 6000 Los Angeles (LA) abrasion cycle groups were analyzed in MATLAB. Particle roundness was quantified using the perimeter-to-area-based Al-Rousan approach[23]. Mean roundness values decreased from 1.494 at 0 cycles to 1.322 at 1000 cycles and 1.240 at 6000 cycles, representing an approximately 17.0% reduction between the initial and final abrasion levels. Based on the Al-Rousan classification, these values correspond to Angular, Subrounded, and Well Rounded aggregate particles, respectively.

This trend may be attributed to the progressive abrasion of particle corners and protrusions, resulting in smoother and more regular particle boundaries. The reduction was more pronounced between 0 and 1000 cycles (0.172) than between 1000 and 6000 cycles (0.082), suggesting that the most prominent geometric irregularities were preferentially removed during the initial abrasion stage. These findings indicate that LA abrasion is associated with changes in particle morphology in addition to mass loss. Image-based roundness measurements therefore provide complementary information to conventional LA abrasion loss measurements, consistent with the quantitative aggregate shape characterization approach described by Al-Rousan[23].

Overall, increasing LA abrasion cycles were associated with more regular aggregate particle geometries and a progressive shift toward Well Rounded shapes. These morphological differences provided the basis for evaluating the classification performance of six CNN architectures in distinguishing Angular, Subrounded, and Well Rounded aggregate particles.

3.2. CNN model training and independent test performance

The training and independent test performance of six CNN architectures—DenseNet-201, MobileNet-V2, ResNet-18, ResNet-50, ResNet-101, and SqueezeNet—was comparatively evaluated. Previous studies have demonstrated the applicability of CNN-based approaches to automated particle shape characterization and aggregate image classification[27,28]. The training accuracy and loss curves are presented in Fig. **3.** All models exhibited increasing accuracy and decreasing loss during training, indicating progressive optimization. DenseNet-201, MobileNet-V2, ResNet-18, ResNet-50, and ResNet-101 showed relatively rapid convergence, whereas SqueezeNet exhibited slower initial convergence. Validation accuracies ranged from 95.50% for SqueezeNet to 99.44% for ResNet-101. Despite these high validation accuracies, independent test performance was evaluated separately to assess model generalization.

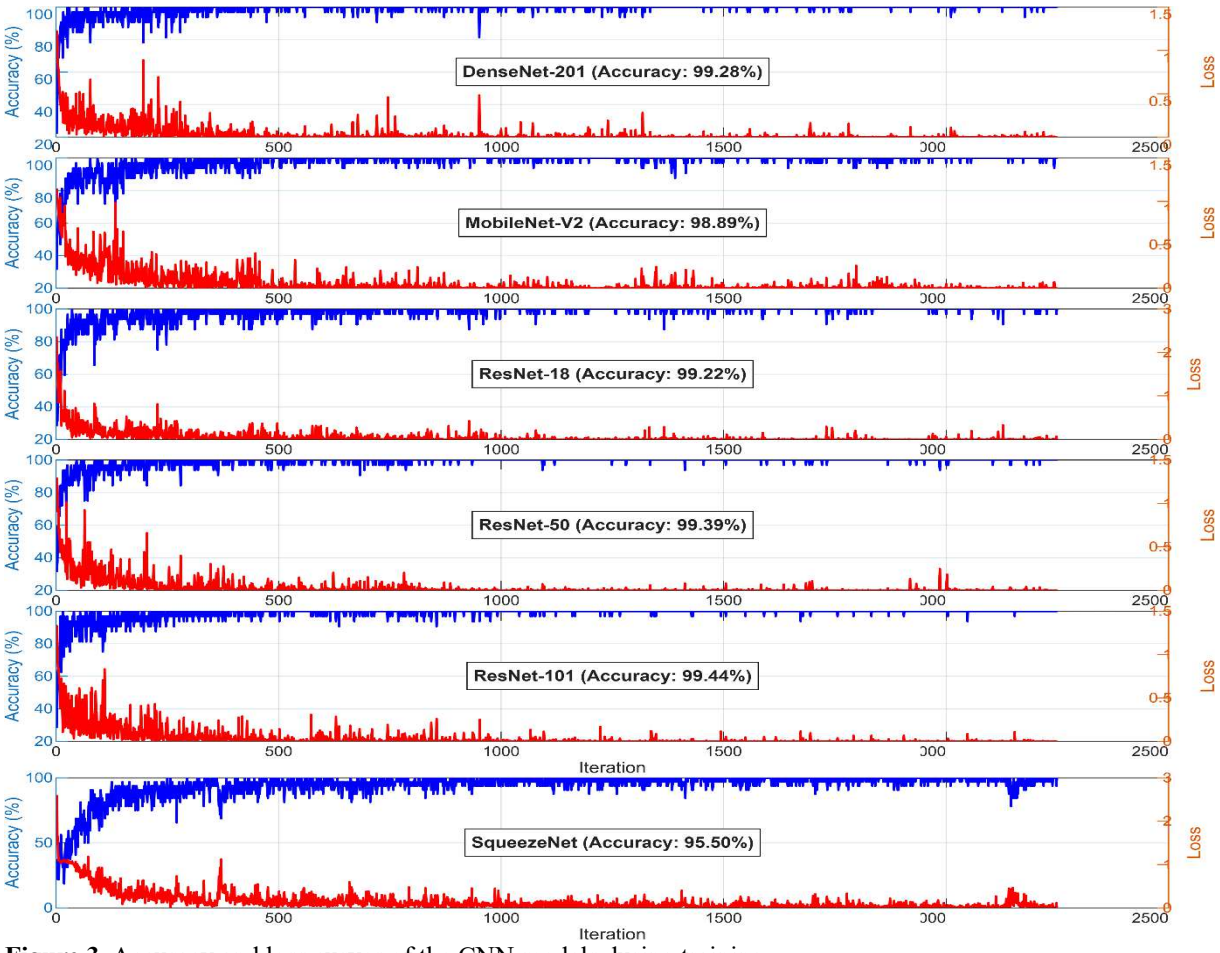


Figure 3. Accuracy and loss curves of the CNN models during training.

ResNet-101 achieved the highest validation accuracy (99.44%), followed by ResNet-50 (99.39%), DenseNet-201 (99.28%), ResNet-18 (99.22%), MobileNet-V2 (98.89%), and SqueezeNet (95.50%). These results indicate that all six architectures effectively learned discriminative visual features associated with the angular, subrounded, and well rounded classes. However, high validation accuracy does not necessarily indicate strong generalization to unseen images, highlighting the importance of independent testing in image-based particle classification[29,30].

Therefore, all models were further evaluated using 300 independent test images excluded from training and validation. The results are presented in Table 1**.**

Table 1. Training and independent test performance of the CNN models

Model	Validation Accuracy (%)	Independent Test Accuracy (%)
DenseNet-201	99.28	91.33
MobileNet-V2	98.89	64.67
ResNet-18	99.22	64.00
ResNet-50	99.39	82.00
ResNet-101	99.44	79.00
SqueezeNet	95.50	78.00

DenseNet-201 achieved the highest independent test accuracy (91.33%), followed by ResNet-50 (82.00%), ResNet-101 (79.00%), SqueezeNet (78.00%), MobileNet-V2 (64.67%), and ResNet-18 (64.00%). Notably, ResNet-101 achieved the highest validation accuracy (99.44%), whereas DenseNet-201 performed best on the independent test dataset. This discrepancy highlights that high validation accuracy does not necessarily translate into superior generalization to unseen images. Previous studies have similarly reported that image-based particle classification performance may vary with dataset characteristics and particle morphology[28]. Although DenseNet-201 exhibited a decrease in accuracy from 99.28% during validation to 91.33% on the independent test dataset, it demonstrated the strongest generalization performance among the six architectures evaluated. In contrast, ResNet-18 achieved the lowest test accuracy (64.00%). As overall accuracy does not reveal class-specific misclassification patterns, the predictions for the angular, subrounded, and well rounded classes were further examined using confusion matrices.

3.3. Class-Specific prediction performance and error analysis

Overall accuracy alone does not fully characterize class-specific performance or misclassification patterns. Therefore, the predictions of the six CNN architectures were further examined using confusion matrices, which compare true and predicted classes and identify interclass classification errors[31,32]. Previous studies have also emphasized the importance of class-specific evaluation beyond overall accuracy[33,34]. The analysis was conducted on 300 independent test images (100 per class), comprising Angular, Subrounded, and Well Rounded aggregate particles. The resulting confusion matrices are presented in Fig. 4.

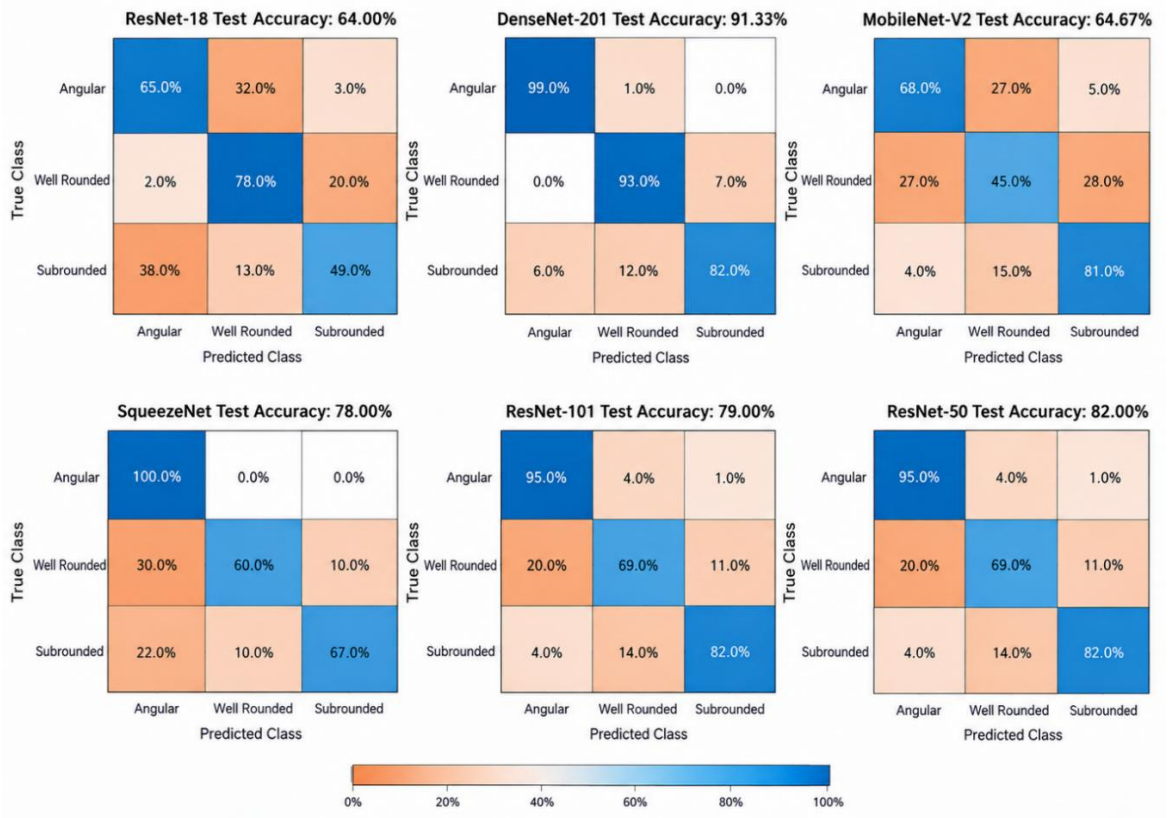


Figure 4. Class-specific prediction results and confusion matrices of the CNN models on the independent test dataset.

As shown in Fig. 4**,** the investigated models exhibited different levels of performance in distinguishing among the three roundness classes. In addition to overall accuracy, Precision, Recall, and F1-score values were calculated to evaluate classification performance in greater detail.

For the Well Rounded class, Recall values of 69%, 78%, 45%, 93%, 60%, and 71% were obtained for ResNet-50, ResNet-18, MobileNet-V2, DenseNet-201, SqueezeNet, and ResNet-101, respectively. Recall represents the proportion of actual samples belonging to a given class that are correctly assigned to that class. DenseNet-201 achieved the highest Recall (93%), followed by ResNet-18 (78%), whereas MobileNet-V2 exhibited the lowest value (45%). These differences indicate that the recognition of Well Rounded particles varied substantially depending on the CNN architecture.

For the Angular class, SqueezeNet achieved the highest Recall (100%), followed by DenseNet-201 and ResNet-101 (99%), ResNet-50 (95%), MobileNet-V2 (68%), and ResNet-18 (65%). Precision represents the proportion of samples predicted as belonging to a particular class that genuinely belong to that class. DenseNet-201 achieved the highest Precision (94.29%), followed by ResNet-50 (79.83%), ResNet-101 (74.44%), MobileNet-V2 (68.69%), SqueezeNet (65.79%), and ResNet-18 (61.90%). Although SqueezeNet correctly identified all actual Angular samples, its lower Precision indicates that some particles from other classes were incorrectly predicted as Angular. In contrast, DenseNet-201 combined high Precision and Recall, indicating more balanced recognition of Angular particles. ResNet-18 exhibited comparatively lower performance in both metrics.

For the Subrounded class, Recall values ranged from 49.00% to 82.00%, whereas Precision values varied between 68.06% and 92.13%. DenseNet-201 achieved the highest Precision (92.13%), followed by SqueezeNet (88.10%), ResNet-50 (87.23%), ResNet-101 (78.82%), MobileNet-V2 (71.05%), and ResNet-18 (68.06%). DenseNet-201 and ResNet-50 achieved the highest Recall (82.00%), followed by MobileNet-V2 (81.00%), SqueezeNet (74.00%), ResNet-101 (67.00%), and ResNet-18 (49.00%). The particularly low Recall obtained by ResNet-18 indicates that a substantial proportion of actual Subrounded samples were incorrectly assigned to other classes. In contrast, DenseNet-201 combined high Precision and Recall, demonstrating stronger recognition of this morphological category. The transitional geometric characteristics of Subrounded particles between the Angular and Well Rounded classes may have contributed to the classification difficulty observed for some architectures.

Overall, the class-specific results revealed substantial differences in classification performance among the investigated CNN architectures. DenseNet-201 demonstrated consistently strong recognition performance across the three morphological classes, whereas ResNet-18 exhibited greater difficulty in identifying Subrounded particles. These findings indicate that class-specific Precision, Recall, and F1-score values provide important insights into classification behavior that cannot be obtained from overall accuracy alone.

3.4. Comparison of the models and evaluation of independent test performance

The performances of the six CNN architectures were compared by jointly considering validation results, independent test accuracy, and class-specific prediction metrics. Performance on the independent test dataset was used as the primary criterion for assessing model generalization. Although all models achieved high validation accuracies, their performance on previously unseen aggregate images varied considerably. The independent test results are summarized in Table 2.

Table 2. Comparison of the overall performance of the CNN models on the independent test dataset.

Model	Accuracy (%)	Macro Precision (%)	Macro Recall (%)	Macro F1 (%)
DenseNet-201	91.33	91.39	91.33	91.22
ResNet-50	82.00	82.13	82.00	81.70
ResNet-101	79.00	79.95	79.00	78.48
SqueezeNet	78.00	82.54	78.00	77.66
MobileNet-V2	64.67	63.82	64.67	64.06
ResNet-18	64.00	64.46	64.00	63.45

As presented in Table 2, DenseNet-201 demonstrated the highest independent test accuracy among the six CNN architectures, whereas ResNet-18 exhibited the lowest performance. The differences in test accuracy indicate that the investigated architectures varied considerably in their ability to classify previously unseen aggregate images. These findings further highlight the importance of evaluating model performance using independent test data rather than relying solely on validation accuracy.

To further evaluate classification performance, Macro Precision, Macro Recall, and Macro F1-score were examined alongside overall accuracy. These metrics represent the unweighted averages of the corresponding class-specific Precision, Recall, and F1-score values, thereby providing a balanced assessment across the three morphological categories. DenseNet-201 achieved the highest Macro Precision (91.39%), Macro Recall (91.33%), and Macro F1-score (91.22%). ResNet-50 ranked second in terms of accuracy and Macro F1-score, achieving 82.00% and 81.70%, respectively. The remaining architectures exhibited lower Macro F1-scores, ranging from

63.45% for ResNet-18 to 78.48% for ResNet-101. These findings further support the superior overall classification performance of DenseNet-201 on the independent test dataset.

The class-specific F1-score results revealed additional differences in the ability of the models to distinguish among the Angular, Subrounded, and Well Rounded categories. DenseNet-201 achieved F1-scores of 96.59%, 90.29%, and 86.77% for the Angular, Well Rounded, and Subrounded classes, respectively. The corresponding values were 86.76%, 73.80%, and 84.54% for ResNet-50; 84.98%, 78.02%, and 72.43% for ResNet-101; 79.37%, 73.17%, and 80.43% for SqueezeNet; 68.34%, 48.13%, and 75.70% for MobileNet-V2; and 63.41%, 69.96%, and 56.98% for ResNet-18. The comparatively low F1-score of MobileNet-V2 for the Well Rounded class (48.13%) and that of ResNet-18 for the Subrounded class (56.98%) indicate that classification difficulty varied across both architectures and morphological categories. The transitional geometric characteristics of Subrounded particles may have contributed to misclassification, while the recognition of Well Rounded particles also presented challenges for certain models.

Overall, DenseNet-201 demonstrated the most consistent classification performance among the six architectures, achieving the highest independent test accuracy and Macro F1-score. Nevertheless, the differences observed across individual morphological classes indicate that overall accuracy alone is insufficient to fully characterize model performance. Accordingly, the combined evaluation of Accuracy, Precision, Recall, F1-score, and class-specific misclassification patterns provides a more comprehensive framework for assessing CNN-based aggregate roundness classification[34,35].

3.5. Pairwise comparison of the models using the McNemar test

Pairwise McNemar tests were performed to compare the classification performance of the six CNN models on the same 300 independent test images. For each model pair, discordant predictions were identified as images correctly classified by the first model but misclassified by the second (b), and vice versa (c). All 15 possible pairwise comparisons were conducted, and the resulting p-values were adjusted using the Holm correction to account for multiple testing. The results are presented in Fig. 5.

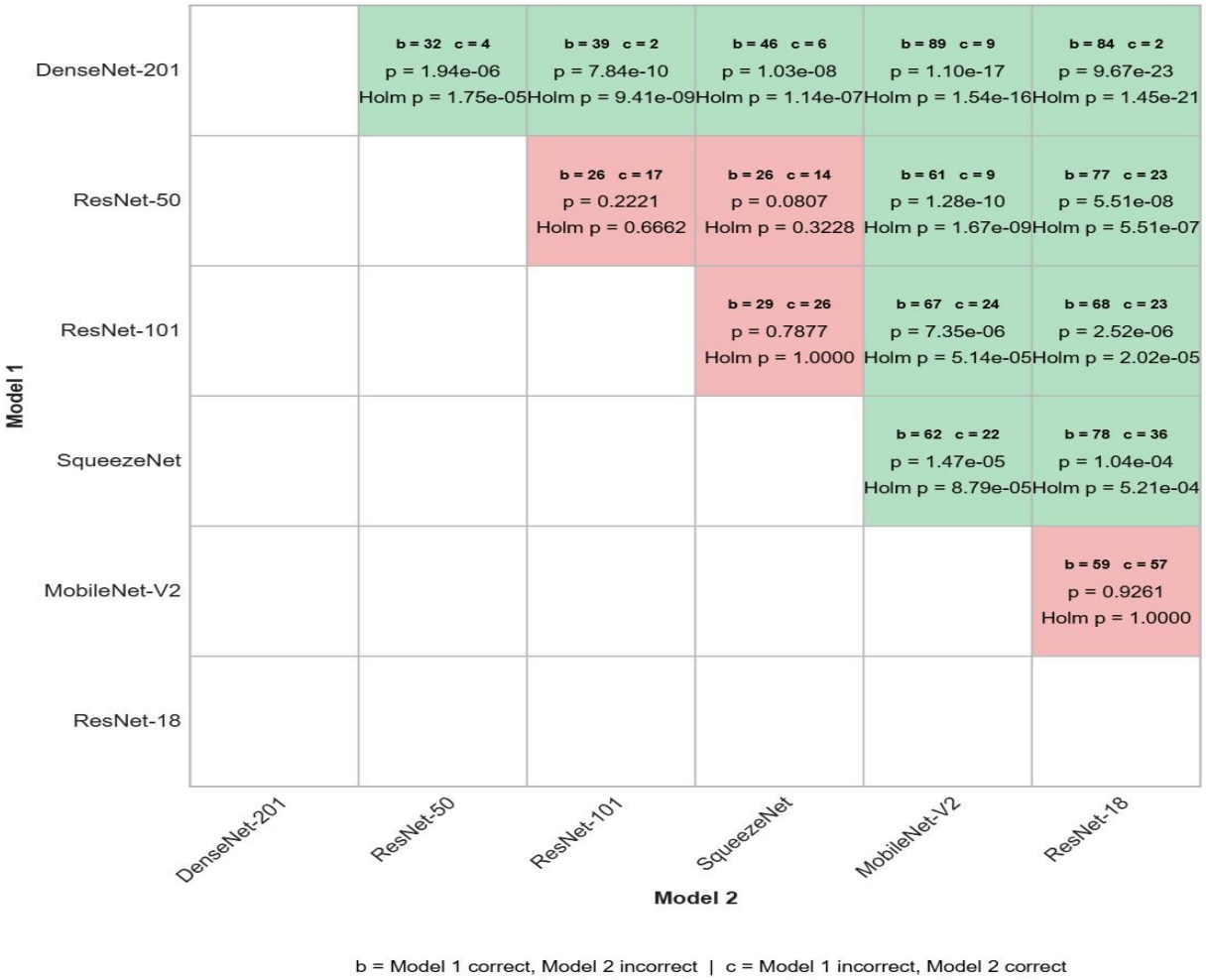


Figure 5. Pairwise McNemar test results of the CNN models on the independent test dataset.

Note: Green cells indicate statistically significant model comparisons with $p < 0.05$ after Holm correction, whereas red cells represent comparisons with $p \geq 0.05$. *b* denotes the number of images correctly classified by Model 1 but

incorrectly classified by Model 2, while *c* represents the number of images incorrectly classified by Model 1 but correctly classified by Model 2. Both the raw *p*-value and the Holm-adjusted *p*-value are reported within each cell.

As shown in Fig. 5, statistically significant differences were identified in 11 of the 15 pairwise comparisons after Holm correction (adjusted $p < 0.05$). DenseNet-201 differed significantly from all five other models, whereas no significant differences were observed between ResNet-50 and ResNet-101, ResNet-50 and SqueezeNet, ResNet-101 and SqueezeNet, or MobileNet-V2 and ResNet-18. The strongest differences involving DenseNet-201 were observed against ResNet-50 (b = 32, c = 4; adjusted $p = 1.75 \times 10^{-5}$) and ResNet-18 (b = 84, c = 2; adjusted $p = 1.45 \times 10^{-21}$), indicating significantly more correct predictions by DenseNet-201 on the same independent test images. In contrast, the non-significant comparisons showed more balanced discordant outcomes, including ResNet-50 versus ResNet-101 (b = 26, c = 17), ResNet-50 versus SqueezeNet (b = 26, c = 14), ResNet-101 versus SqueezeNet (b = 29, c = 26), and MobileNet-V2 versus ResNet-18 (b = 59, c = 57; adjusted $p = 1.0000$). These findings demonstrate that the observed differences in overall test accuracy were not statistically significant for every model pair.

3.6. Grad-CAM visualization of the DenseNet-201 model

Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to DenseNet-201 to visualize the image regions contributing to its classification decisions. Grad-CAM visualizations were generated for correctly classified images from the independent test dataset, representing the Angular, Subrounded, and Well Rounded classes. Fig. 6 presents the resulting activation maps alongside the corresponding original aggregate images.

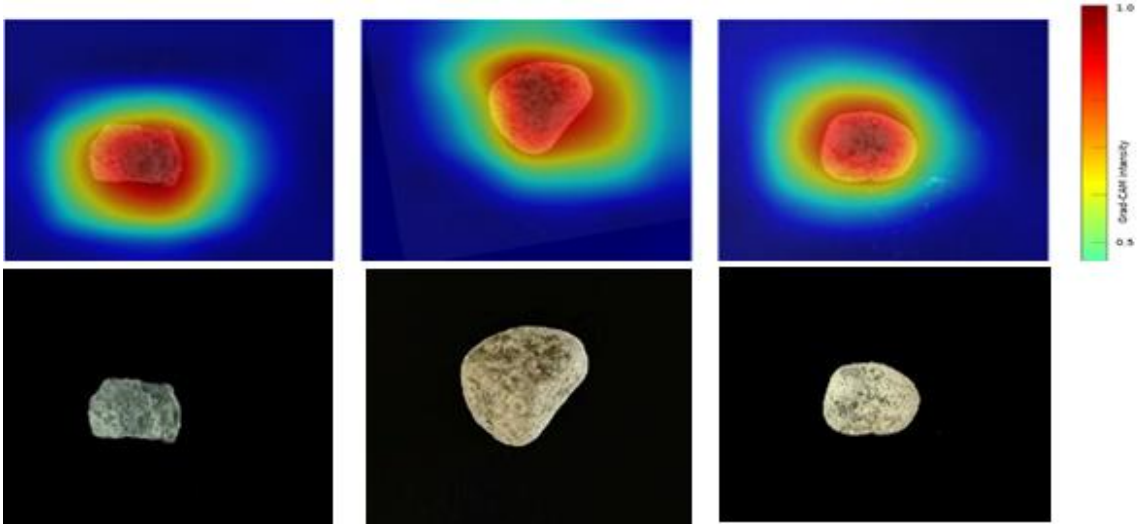


Figure 6. Grad-CAM visualization of the regions influencing the classification decisions of the DenseNet-201 model.

As shown in Fig. 6, DenseNet-201 exhibited high activation primarily within aggregate regions, particularly along particle boundaries and distinctive geometric features. These findings suggest that the model's classification decisions were influenced predominantly by aggregate morphology rather than background features, supporting the relevance of the learned features across the three roundness classes.

4. Conclusions

This study investigated the automated classification of aggregate particle roundness using digital image analysis and six convolutional neural network (CNN) architectures. Aggregate particles subjected to different Los Angeles abrasion cycles were characterized using the Al-Rousan shape factor and classified into Angular, Subrounded, and Well Rounded categories. The mean shape factor decreased from 1.494 at 0 cycles to 1.322 at 1000 cycles and 1.240 at 6000 cycles, indicating progressively more regular particle contours with increasing abrasion. Among the six CNN architectures, DenseNet-201 achieved the highest independent test accuracy (91.33%), followed by ResNet-50 (82.00%), ResNet-101 (79.00%), SqueezeNet (78.00%), MobileNet-V2 (64.67%), and ResNet-18 (64.00%). These findings demonstrate the effectiveness of DenseNet-201 under the investigated experimental conditions, although its performance should not be generalized to other datasets without further validation. Class-specific results revealed greater classification difficulty for Subrounded particles, potentially reflecting their transitional morphology between Angular and Well Rounded particles.

Overall, the findings support the feasibility of image-based deep learning for automated aggregate roundness classification. However, the study was limited to aggregates from a single source and the investigated imaging conditions. Future research should evaluate the approach using larger and more diverse datasets encompassing

different lithologies, particle sizes, aggregate sources, and imaging conditions. Multi-view or three-dimensional imaging could further improve morphological characterization. Additionally, combining image-derived morphological descriptors with experimental measurements may enable the development of predictive models for engineering properties such as compaction and California Bearing Ratio (CBR). Such models could complement conventional laboratory testing by supporting preliminary assessment and rapid prediction.

Declarations

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This study was supported by the Atatürk University Scientific Research Projects Coordination Unit (BAP) under Project No. 16034.

Author contributions

All authors contributed to the conceptualization of the study and the development of its methodological approach. Gokhan Ucar and Fatih Isık were responsible for the conceptualization of the study and the development of the methodology. Gokhan Ucar carried out data curation, software development, investigation, formal analysis, and visualization. Fatih Isık was responsible for providing resources, validation, supervision, and project administration. Gokhan Ucar and Fatih Isık contributed to the preparation of the original draft of the manuscript. Fatih Isık was responsible for reviewing and editing the manuscript. All authors reviewed the results and read and approved the final version of the manuscript.

Acknowledgements

Not applicable.

Authors' information

Not applicable.

Declaration of generative AI in scientific writing

This study used ChatGPT to assist with translation and code review as part of the research process.

References

1. Krumbein, W. C. & Sloss, L. L. *Stratigraphy and Sedimentation*. 2nd edn. (W. H. Freeman and Company, San Francisco, CA, USA, 1963).
2. Pan, T., Tutumluer, E. & Anochie-Boateng, J. Aggregate morphology affecting resilient behavior of unbound granular materials. *Transp. Res. Rec.* **1952**, 12–20 (2006). https://doi.org/10.1177/0361198106195200102
3. Ali, U., Kikumoto, M., Ciantia, M., Cui, Y. & Previtali, M. Systematic effect of particle roundness/angularity on macro- and microscopic behavior of granular materials. *Granular Matter* **25**, 51 (2023). https://doi.org/10.1007/s10035-023-01341-y
4. Fletcher, T., Chandan, C., Masad, E. & Sivakumar, K. Aggregate imaging system for characterizing the shape of fine and coarse aggregates. *Transp. Res. Rec.* **1832**, 67–77 (2003). https://doi.org/10.3141/1832-09
5. Masad, E. & Button, J. W. Unified imaging approach for measuring aggregate angularity and texture. *Comput.-aided Civ. Infrastruct. Eng.* **15**, 273–280 (2000). https://doi.org/10.1111/0885-9507.00191
6. Al-Rousan, T., Masad, E., Tutumluer, E. & Pan, T. Evaluation of image analysis techniques for quantifying aggregate shape characteristics. *Constr. Build. Mater.* **21**, 978–990 (2007). https://doi.org/10.1016/j.conbuildmat.2006.03.005
7. Chandan, C., Sivakumar, K., Masad, E. & Fletcher, T. Application of imaging techniques to geometry analysis of aggregate particles. *J. Comput. Civ. Eng.* **18**, 75–82 (2004). https://doi.org/10.1061/(ASCE)0887-3801(2004)18:1(75)

8. Lee, J. R. J., Smith, M. L. & Smith, L. N. A new approach to the three-dimensional quantification of angularity using image analysis of the size and form of coarse aggregates. *Eng. Geol.* **91**, 254–264 (2007). https://doi.org/10.1016/j.enggeo.2007.02.003
9. Huang, H., Luo, J., Tutumluer, E., Hart, J. M. & Stolba, A. J. Automated segmentation and morphological analyses of stockpile aggregate images using deep convolutional neural networks. *Transp. Res. Rec.* **2674**, 285–298 (2020). https://doi.org/10.1177/0361198120943887
10. Yaseen, G., Alaloul, W. S., Hafeez, I. & Qureshi, A. H. Shape characterizing of aggregates produced through different crushing techniques. *Coatings* **11**, 1199 (2021). https://doi.org/10.3390/coatings11101199
11. Zhao, X.-C. *et al.* Computer vision technology for characterizing particle size and shape of aggregate materials: A review. in *CICTP 2021* 906–916 (American Society of Civil Engineers, Reston, VA, 2021). https://doi.org/10.1061/9780784483565.087
12. Ma, W., Han, T., Xu, Z. & Lin, P. Feature fusion of single and orthogonal polarized rock images for intelligent lithology identification. *AI Civil Eng.* **4**, (2025). https://doi.org/10.1007/s43503-025-00049-7
13. Lateef, J. & Yu, X. Novel computer vision algorithm for accurate multi-label classification of concrete bridge defects. *AI Civil Eng.* **5**, 10 (2026). https://doi.org/10.1007/s43503-026-00095-9
14. Hassan, R., Onyelowe, K. C. & Zamel, A. A. Image processing and neural network technique for size characterization of gravel particles. *Sci. Rep.* **14**, 22737 (2024). https://doi.org/10.1038/s41598-024-72700-9
15. Shoji, D., Noguchi, R., Otsuki, S. & Hino, H. Classification of volcanic ash particles using a convolutional neural network and probability. *Sci. Rep.* **8**, 8111 (2018). https://doi.org/10.1038/s41598-018-26200-2
16. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015). https://doi.org/10.1038/nature14539
17. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (IEEE, 2016). http://doi.org/10.1109/CVPR.2016.90
18. Zhang, P., Yin, Z.-Y. & Jin, Y.-F. Three-dimensional quantitative analysis on granular particle shape using convolutional neural network. *Int. J. Numer. Anal. Methods Geomech.* **46**, 187–204 (2022). https://doi.org/10.1002/nag.3296
19. Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0
20. McNemar, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* **12**, 153–157 (1947). https://doi.org/10.1007/BF02295996
21. Holm, S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65–70 (1979). https://www.jstor.org/stable/4615733
22. ASTM International. *ASTM C131/C131M-20: Standard Test Method for Resistance to Degradation of Small-Size Coarse Aggregate by Abrasion and Impact in the Los Angeles Machine* (ASTM International, 2020). http://doi.org/10.1520/C0131_C0131M-20
23. Al-Rousan, T. M. *Characterization of Aggregate Shape Properties Using a Computer Automated System*. (College Station, TX, USA, 2004).
24. Arasan, S., Akbulut, S. & Hasiloglu, A. S. The relationship between the fractal dimension and shape properties of particles. *KSCE J. Civ. Eng.* **15**, 1219–1225 (2011). https://doi.org/10.1007/s12205-011-1310-x
25. Selvaraju, R. R. *et al.* Grad-CAM: Visual explanations from deep networks via gradient-based localization. in *2017 IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2017). http://doi.org/10.1109/ICCV.2017.74
26. Hasan, M. & Lu, M. Bridging AI and explainability in civil engineering: the Yin-Yang of predictive power and interpretability. *AI Civil Eng.* **4**, (2025). https://doi.org/10.1007/s43503-025-00066-6
27. Chen, Z. *et al.* Image-based methods for automatic identification of elongated and flat aggregate particles. *Constr. Build. Mater.* **382**, 131187 (2023). http://doi.org/10.1016/j.conbuildmat.2023.131187
28. Subramaniam, D. N., Sajeevan, M., Pratheeba, J., Wijekoon, S. H. B. & Sathiparan, N. Characterization of the shape of aggregates using image analysis and machine learning classification tools. *Geomech. Geoengin.* **19**, 421–443 (2024). http://doi.org/10.1080/17486025.2023.2288655
29. Zheng, Y., Huang, J., Chen, T., Ou, Y. & Zhou, W. Transfer of learning in the convolutional neural networks on classifying geometric shapes based on local or global invariants. *Front. Comput. Neurosci.* **15**, 637144 (2021). http://doi.org/10.3389/fncom.2021.637144
30. Bals, J. & Epple, M. Deep learning for automated size and shape analysis of nanoparticles in scanning electron microscopy. *RSC Adv.* **13**, 2795–2802 (2023). http://doi.org/10.1039/D2RA07812K
31. Congalton, R. G. A review of assessing the accuracy of classification of remotely sensed data. *Remote Sens. Environ.* **37**, 35–46 (1991). http://doi.org/10.1016/0034-4257(91)90048-B
32. Stehman, S. V. Selecting and interpreting measures of thematic classification accuracy. *Remote Sens. Environ.* **62**, 77–89 (1997). http://doi.org/10.1016/S0034-4257(97)00083-7
33. Foody, G. M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **80**, 185–201 (2002). http://doi.org/10.1016/S0034-4257(01)00295-4
34. Sokolova, M. & Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **45**, 427–437 (2009). http://doi.org/10.1016/j.ipm.2009.03.002
35. Powers, D. M. W. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.* **2**, 37–63 (2011). http://doi.org/10.9735/2229-3981