JOURNAL of
MATERIALS and MECHATRONICS:A

e-ISSN 2717-8811
JournalMM, 2025, 6(2), 442-462
https://doi.org/10.55546/jmm.1769352

Araştırma Makalesi / Research Article

Superpixel-Based Image Analysis for Stripping Detection in Asphalt Mixtures with ANN and U-Net Models

Kadir AKGÖL[1*], Mehmet Can TUNA[2]

[1*] Giresun Üniversitesi, Mühendislik Fakültesi, İnşaat Mühendisliği Bölümü, Giresun, Türkiye,
ORCID ID: https://orcid.org/0000-0002-1939-6717, kadir.akgol@giresun.edu.tr
[2*] Giresun Üniversitesi, Mühendislik Fakültesi, İnşaat Mühendisliği Bölümü, Giresun, Türkiye,
ORCID ID: https://orcid.org/0009-0007-9368-4155, mehmetcantunaa@hotmail.com

Geliş/ Received: 20.08.2025; **Revize/Revised:** 09.12.2025 **Kabul / Accepted:** 18.12.2025

ABSTRACT: The reliable detection of stripping in asphalt mixtures is a critical challenge for pavement performance evaluation, as conventional physical tests rely heavily on subjective observation and lack reproducibility. This study proposes an image-based quantitative method that integrates geometric standardization, superpixel segmentation, and feature extraction to enhance the objectivity of stripping assessment. Petri dish images were first standardized through square cropping and bicubic resampling to ensure comparability across samples. Superpixels were then generated, and multiple spatial, geometric, photometric, and texture-based features were extracted, including distance-to-center, compactness, local color similarity, and global color deviation. Automatic background labeling was achieved through a color-based masking approach validated by visual inspection. The extracted feature set was subsequently employed for supervised classification using artificial neural networks (ANNs), with model performance evaluated against reference segmentations. The results demonstrate that the proposed method achieves high classification accuracy, with robust generalization across multiple sample sets. In particular, ANN-based predictions exhibited superior discriminative capability in distinguishing stripped from coated aggregate regions, outperforming U-Net segmentation under identical input conditions. The findings highlight that incorporating contextual descriptors, such as black pixel ratio and blue-background masking, significantly improves classification robustness in low-contrast and noisy regions. Overall, the proposed framework provides a reproducible and efficient alternative to conventional stripping tests, enabling reliable quantitative evaluation of asphalt mixture performance. This study contributes to the advancement of automated image analysis methods in pavement engineering and establishes a foundation for broader integration of computer vision into asphalt durability assessment.

Keywords: Stripping detection, Image analysis, Superpixel segmentation, Artificial neural networks

*Sorumlu yazar / Corresponding author: kadir.akgol@giresun.edu.tr
Bu makaleye atıf yapmak için /To cite this article

Akgöl, K., Tuna, M. C. (2025). Superpixel-Based Image Analysis for Stripping Detection in Asphalt Mixtures with ANN and U-Net Models. Journal of Materials and Mechatronics: A (JournalMM), 6(2), 442-462.

1. INTRODUCTION

The durability and service life of asphalt pavements largely depend on the quality of adhesion between the aggregate and the bitumen binder. Weak bonding can lead to the initiation of stripping on the surface over time. This allows water to penetrate between layers and causes damage in the underlying courses. Ultimately, such damage jeopardizes the structural integrity of the pavement. Surface dressing systems are widely applied because of their low cost. However, the failure to detect such degradations at an early stage not only adversely affects structural performance, but also substantially increases maintenance and repair costs. Therefore, the reliable evaluation of the aggregate–bitumen interface interaction is of critical importance both for engineering applications and for academic studies.

The evaluation of stripping resistance in asphalt mixtures is generally carried out through laboratory-based physical methods such as boiling tests, freeze-thaw cycles, and moisture sensitivity tests. These methods aim to assess adhesion resistance by simulating adverse environmental conditions to which the pavement material will be exposed during its service life. For example, in the study by Öner (2020), the stripping resistance of asphalt mixtures prepared with different proportions of granite ceramic waste was examined using the Nicholson Stripping Test, and it was determined that granite ceramic waste at a rate of 20% could replace traditional limestone. In such studies in the literature, visual-based quantitative analyses are not included, and evaluations are often carried out based on observation. Because stripping detection processes rely on expert judgment, differences in interpretation may arise from one user to another. Consequently, this situation hinders the standardization of the obtained results. Moreover, the fact that these physical methods are both time-consuming and limited in terms of repeatability increasingly highlights the need for digital and automated methods.

In this context, some recent studies have turned to computer-based image processing techniques in order to make stripping detection independent of expert opinion, objective, and repeatable. Xiao, Polaczyk, and Huang (2022) developed color-based segmentation methods to automatically determine the stripping ratio after the boiling test. Similarly, M. Li et al. (2023) performed feature extraction in different color spaces and applied classification algorithms to facilitate the distinction between bitumen-coated and stripped aggregate regions. Cui, Wu, Xiao, Wang, and Wang (2019) proposed a preprocessing step that normalized lighting conditions to minimize color variations in images and then performed segmentation based on the color components of pixels. In the study by Gürer and Karaşahin (2016), a color threshold-based algorithm was used to determine the stripping percentage from boiling test images. The applicability of image-processing-based approaches to stripping detection has been demonstrated by these studies. However, the methods employed were generally restricted to the color properties of pixels. In these studies, neither contour information nor a generalizable background-separation strategy robust to data diversity was developed.

Image-based techniques have also been widely used to characterize aggregate particles in terms of shape, angularity, surface texture, and related morphology. Several studies extracted size and shape descriptors from high-resolution images to investigate the influence of crushing and production processes on aggregate morphology and mixture performance (Kamani & Ajalloeian, 2022; Reddy, Abdallah, & Nazarian, 2025; Théodon, Coufort-Saudejaud, Hamieh, & Debayle, 2023; H. N. Wang et al., 2020; L. B. Wang, Lane, Lu, & Druta, 2009). Other works developed image-processing algorithms for the automatic extraction of aggregate geometry and shape factors (Sinecen & Makinaci, 2010; Sinecen, Makinaci, & Topal, 2011). More recently, image-based approaches have also been employed to analyze aggregate distribution, packing density, and gradation in asphalt

mixtures by locating particles, deriving gradation curves, and estimating void ratios or packing density from segmented images (Cao, Zhao, Gao, Huang, & Zhang, 2019; H. H. Huang, Luo, Tutumluer, Hart, & Stolba, 2020; T. Huang & Liu, 2024; Reyes-Ortiz, Mejia, & Useche-Castelblanco, 2021; Salemi & Wang, 2018; Xing, Xu, Tan, Liu, & Ye, 2019). In all of these studies, segmentation constitutes a crucial preliminary step and shares common technical components with the present work; however, their primary focus is on geometric characterization (morphology, gradation, packing) rather than on explicitly quantifying moisture-induced stripping or performing multi-class segmentation tasks such as background separation and stripping ratio determination.

In more recent times, deep learning–based methods and advanced segmentation algorithms have begun to be used to separate complex and overlapping particles. Zong, Zhou, Li, and Wang (2023), using the Mask R-CNN architecture, successfully detected aggregate particles of different shapes and sizes with high accuracy; the model was able to separate even overlapping grains at the contour level. H. J. Li, Asbjörnsson, and Lindqvist (2021) and Yan, Liao, Wu, Xie, and Xia (2021) performed automatic segmentation of concrete and mineral particles using convolutional neural networks (CNN), thereby reducing classification errors particularly in edge regions. Peng, Ying, Kamel, and Wang (2020), on the other hand, combined multi-stage segmentation and edge enhancement steps in complex mineral grains, extracting the geometric properties of the particles more reliably. In the study by H. H. Huang et al. (2020), the measurement accuracy was also improved through a deep learning–assisted separation process. The common point of these studies is that they provide high-accuracy object segmentation; however, the fact that the labeling process requires intensive labor and that their focus is directed toward general particle separation rather than background–object distinction in the context of stripping limits the direct applicability of the methods.

Although numerous studies in the literature have employed image processing or learning-based methods for stripping detection in asphalt specimens, a substantial portion of these studies has either not addressed the reliable separation of the background at all or has been limited to rudimentary techniques. Most of the existing approaches have either focused on holistic shape criteria without achieving pixel-level discrimination or have lacked noise-resistant steps such as superpixel-supported segmentation. However, in Petri dish images, the background is not merely a passive region occupying a large portion of the image but also a critical reference point for the accurate classification of bitumen-coated and stripped aggregate areas. Reflections from the transparent surface of the dish, variable lighting, color variations, and bitumen splashes complicate this process and restrict the generalizability of existing segmentation algorithms. In traditional physical methods, since the specimen is directly examined, the results may vary depending on the observer, preventing standardization.

In this context, the primary problem addressed in this study is the development of a segmentation approach that is resistant to visual noise, robust against sample variability, and specifically oriented toward background detection. The aim of the study is to design a learning-based yet highly interpretable solution architecture that enables accurate and generalizable identification of the background in Petri dish images. To this end, the images were standardized in terms of size and framing, expert annotations were generated through a blue-referenced labeling method, and the performance of two distinct learning-based models (a feature-based artificial neural network and U-Net) was compared. Feature importance analyses were conducted to examine the decision-making mechanism of the model, and the proposed method was demonstrated to be strong in terms of both accuracy and interpretability.

2. MATERIALS AND METHODS

The method developed in this study consists of two main stages. In the first stage, the images were transformed into a standardized format in terms of size, framing, and color characteristics to ensure consistency and reproducibility of the analyses. In the second stage, both a feature-based artificial neural network and a U-Net-based deep learning architecture were trained using these standardized data, and their performances were compared. In this way, the generalizability of the method was tested under different conditions of data diversity and image quality. The following section provides a detailed explanation of the first step of this process, namely image standardization.

2.1 Image Standardization

The sample images obtained from stripping tests do not provide a directly comparable dataset due to variations in camera specifications, shooting angles, and similar factors. To transform these images into a reliable dataset, several preprocessing steps were applied. For the dataset to be constructed consistently, all images needed to be geometrically and photometrically aligned to a common reference frame. In this study, each raw image $I \in R^{H\times W\times 3}$ (with height H, width W, and three color channels, respectively) was aligned with the circular boundary of the Petri dish. The region outside the dish was converted into a distinctive reference background color. The image was then resampled into a square patch. In this way, all images were standardized into a uniform format, ensuring comparability across the dataset (Figure 1.a and Figure 1.b).

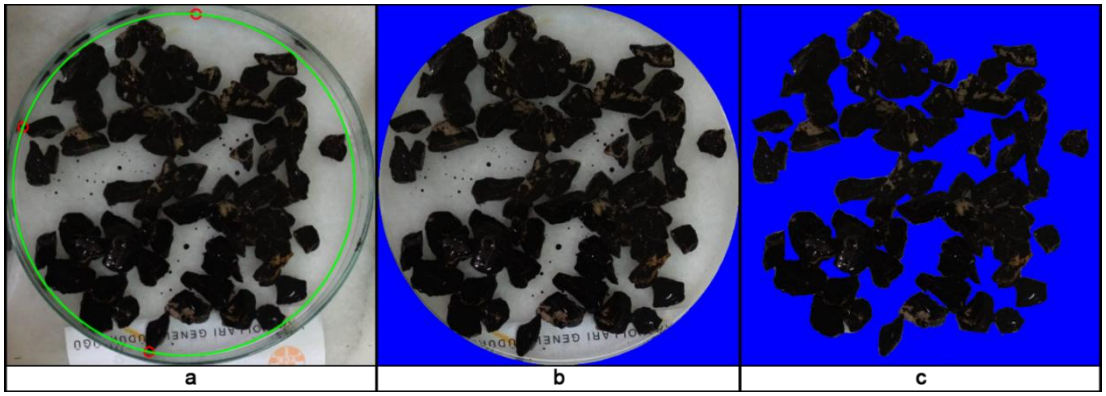

Figure 1. a) Determination of the Petri dish boundary using three selected points (red circles) and visualization of the fitted circle (green circle) on the image, b) generation of the standardized image (I_{std}) with a resolution of 512×512 by masking the outer region of the Petri dish (uniform blue background), c) creation of the reference image (I_{ref}) by coloring the background inside the Petri dish (blue mask)

Let I denote the original RGB image with height H and width W (in pixels). The boundary of the Petri dish is defined by three user-selected points $(x_i, y_i), i = 1, 2, 3$. From these points, the circle parameters, namely the center (x_c, y_c) and radius R, are obtained in closed form. The solution is based on solving the following linear system:

$$A\begin{bmatrix}x_c\\ y_c\end{bmatrix} = B, \qquad A = 2\begin{bmatrix}x_2 - x_1 & y_2 - y_1\\ x_3 - x_1 & y_3 - y_1\end{bmatrix}, \qquad B\begin{bmatrix}x_2^2 - x_1^2 + y_2^2 - y_1^2\\ x_3^2 - x_1^2 + y_3^2 - y_1^2\end{bmatrix} \tag{1}$$

Here, A represents the coefficient matrix constructed from the coordinate differences, while B contains quadratic terms of the point coordinates. Solving this system yields the circle center (x_c, y_c).

Once the center is determined, the radius is calculated the radius R is computed as the Euclidean distance between (x_c, y_c) and any of the three user-defined points (e.g., (x_1, y_1)) (Equation 2).

$$R = \sqrt{(x_c - x_1)^2 + (y_c - y_1)^2} \tag{2}$$

The three-point interaction transfers the Petri dish to the same geometric reference in images captured at different framing and scales, while keeping the user effort low and providing practical robustness against misalignments (Figure 1.a). With the center and radius, the circular region occupied by the dish is defined as the set Ω of pixel coordinates inside the circle:

$$\Omega = \{(x, y) \in \{1, \dots, H\} \times \{1, \dots, H\} | (x - x_c)^2 + (y - y_c)^2 \leq R^2\} \tag{3}$$

where x and y denote integer pixel coordinates and H denotes the image height (and, assuming a square image after cropping, also the width). Over this discrete domain, the binary mask $M(x,y)$ is obtained as

$$M(x, y) = 1_\Omega(x, y) = \begin{cases} 1, (x, y) \in \Omega, \\ 0, otherwise \end{cases} \tag{4}$$

Here, $1_\Omega(x, y)$ denotes the indicator function, taking the value 1 if the pixel (x, y) belongs to the set Ω, and 0 otherwise. Pixels outside the circle (i.e., $M = 0$) are replaced with a uniform blue vector $b = (0,0,255)$, thereby converting the exterior of the dish into a distinct reference background (Figure 1.b). This choice facilitates the subsequent automatic and consistent labeling of the background, since the blue chrominance is distinguishable from the aggregate/bitumen color statistics, enabling the "background" class to be statistically separated from scene-related background candidates (reflections, droplets, transparent edges, etc.) with greater reliability. In the second step of geometric standardization, a square cropping window

$$C = [x_c - R, x_c + R] \times [y_c - R, y_c + R] \tag{5}$$

is defined to fully encompass the Petri dish. In practice, C is intersected with the image boundaries to avoid overflow and ensure that the cropping window remains within the valid image domain. The restriction of the original image I to this window is denoted by $\mathrm{I}|_C$. The resulting patch is then resampled to a resolution of 512×512 using bicubic interpolation,

$$\mathrm{I}_{std} = Resize(\mathrm{I}|_C, 512 \times 512) \tag{6}$$

where Resize(·) denotes the bicubic interpolation operator (Figure 1.b). This resampling procedure ensures the comparability of superpixel scales. It also satisfies the fixed-size input requirements of architectures such as U-Net. Finally, it guarantees fair comparability across samples. Furthermore, manual labeling was performed on copies of the standardized images by coloring the interiors of the Petri dishes in blue, thereby generating reference images I_{ref} (Figure 1.c).

2.2 Feature Extraction and Background Labeling

On the standardized I_{std} images, a superpixel-based segmentation approach was employed to achieve highly accurate background separation by utilizing both geometric and photometric cues. The

primary objective of this step is to partition the pixels into homogeneous subregions, compute statistical features that characterize each region, and assign labels to these regions in a manner suitable for supervised learning. In this way, subsequent classification stages can operate on more meaningful and statistically stable "contour" (superpixel segment) units, rather than on individual pixels.

For the segmentation process, MATLAB's superpixels function was used to divide each image into $N_{sp} = 1000$ superpixels. This number was determined experimentally to ensure sufficient detail in capturing aggregate and bitumen regions while preventing excessive oversegmentation. The resulting superpixel label matrix $S \in \{1, \dots, N_{sp}\}^{512\times512}$ assigns each pixel to its corresponding superpixel identity (Figure 2.a). Subsequently, the boundaries of each superpixel were extracted using the bwboundaries function. These boundary points, denoted as B_k, were indexed and visually inspected on the image for accuracy verification (Figure 2.b). Small noisy regions were removed using a morphological area filter with a threshold of $A_{min} = 50$ pixels, and the preprocessed binary mask of the k-th superpixel was represented as M_k.

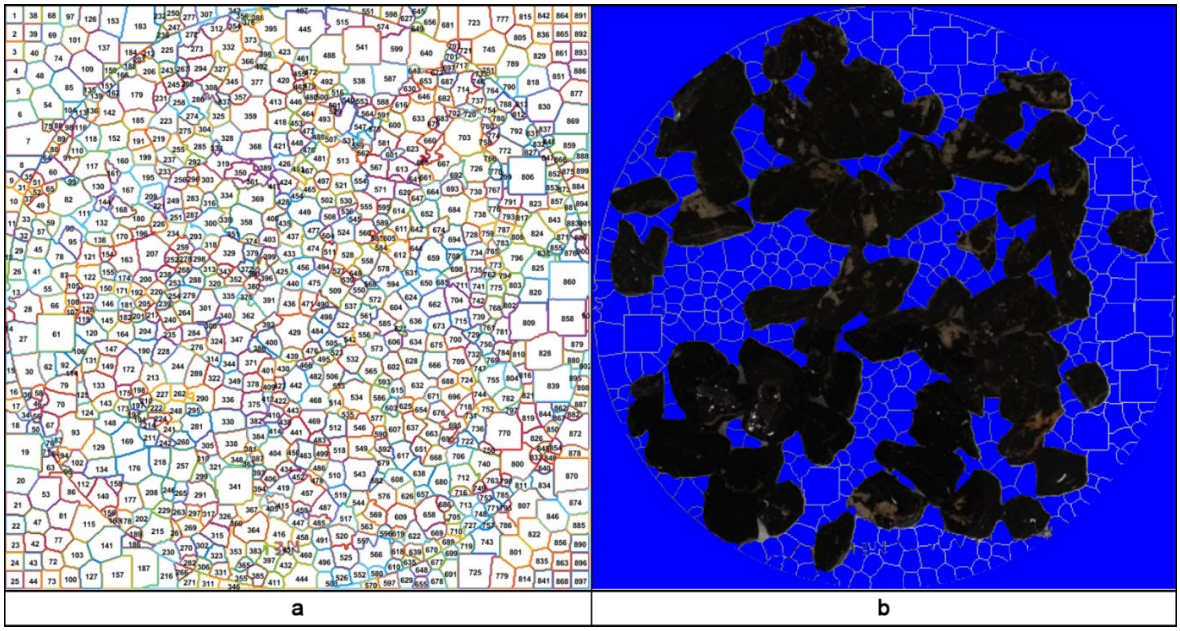

Figure 2. a) Representation of each superpixel with its boundary and unique identifier (ID), b) verification of the alignment of superpixel contours with aggregate edges and background by overlaying them on the original image and the blue background reference

For each superpixel k, spatial, geometric, color, and texture-based features were extracted. The spatial feature was defined as the distance-to-center, computed as the Euclidean distance between the centroid of the superpixel $(c_x^{(k)}, c_y^{(k)})$ and the image center $\left(x_c^{img}, y_c^{img}\right) = (W/2, H/2)$:

$$d_{center}^{(k)} = \sqrt{(c_x^{(k)} - x_c^{img})^2 + (c_y^{(k)} - y_c^{img})^2} \tag{7}$$

Geometric features included the superpixel area A_k (number of pixels within the superpixel mask) and the perimeter P_k (length of the mask boundary), along with the compactness measure:

$$\text{compactness}^{(k)} = \frac{A_k}{P_k^2} \tag{8}$$

Photometric features consisted of the grayscale mean intensity (*meanIntensity*), the median values of each channel in RGB and HSV color spaces $(\tilde{R}_k, \tilde{G}_k, \tilde{B}_k, \tilde{H}_k, S_k, \tilde{V}_k)$, and the grayscale variance $\sigma^2_{gray,k}$. The *meanIntensity* value represents the average brightness of the superpixel, whereas the median values reduce the effect of local outliers such as highlights or shadows, thereby improving statistical robustness.

To capture local color similarities, each superpixel was assigned its $K = 3$ nearest neighbors in terms of radial position (d_{center}). The average ℓ_2-norm distance between the RGB median vectors of the superpixel and its neighbors,

$$\text{neighbor\_color\_diff}^{(k)} = \frac{1}{K} \sum_{j \in N_k} \left\| (\tilde{R}_k, \tilde{G}_k, \tilde{B}_k) - (\tilde{R}_j, \tilde{G}_j, \tilde{B}_j) \right\|_2 \tag{9}$$

was used as a discriminative descriptor, where N_k detones the set of K nearest neighbors of superpixel k and $\|\cdot\|_2$ denotes the Euclidean norm. For the global color context, blue background pixels ($R < 80 \wedge G < 80 \wedge B > 150$) and black pixels ($R < 30 \wedge G < 30 \wedge B < 30$) were excluded, and the mean foreground color was computed as $\bar{f} = (\bar{R}, \bar{G}, \bar{B})$. The distance between the superpixel median RGB vector and this global mean was then calculated as:

$$\text{foreground\_color\_diff}^{(k)} = \left\| (\tilde{R}_k, \tilde{G}_k, \tilde{B}_k) - \bar{f} \right\|_2 \tag{10}$$

and incorporated as an additional feature. The **black_ratio**, defined as the ratio of black pixels over the entire image, was computed once per image and assigned to all superpixels as a constant contextual descriptor. For **automatic background labeling**, the manually painted reference image I_{ref} (Figure 1.c) was used. Within each superpixel mask M_k, the ratio of blue pixels was computed as

$$r^{(k)}_{blue} = \frac{\sum_{(x,y) \in M_k} 1_{blue}(x,y)}{\sum_{(x,y) \in M_k} 1} \tag{11}$$

where $1_{blue}(x,y)$ is a binary indicator function that equals 1 if pixel (x,y) belongs to the "blue" mask, and 0 otherwise. The denominator $\sum_{(x,y) \in M_k} 1$ corresponds to the total number of pixels in M_k. Superpixels with $r^{(k)}_{blue} > \tau_{bg} = 0.5$ were labeled as background, generating the binary label vector $y_{bg} \in \{0,1\}^{N_{sp}}$. In addition to visual inspection of superpixel contours (Figure 2), we quantitatively evaluated the agreement between these superpixel-level labels and the manual background annotations. Across all specimens, the automatic labeling achieved a pixel-wise accuracy of 96.3%, a background precision of 0.975, a recall of 0.950, an F1-score of 0.962, and a background IoU of 0.927, indicating a very high consistency between the superpixel-based labels and the manual reference masks and confirming that the automatic labeling procedure provides a reliable basis for subsequent supervised learning.

As a result, a data table *T* was constructed, where each row corresponds to a superpixel and the columns consist of 15 features (geometric: distance to center, area, perimeter, compactness; photometric: meanIntensity, medianR, medianG, medianB, medianH, medianS, medianV, gray variance; contextual/neighborhood: neighbor color diff, foreground color diff, black ratio) along with

the 'background' label (Table 1). A subset of these features was defined by Equations (7)–(10), while the remaining ones were computed according to standard definitions.

Table 1. Selected features of the contours obtained from the image processing procedure

Dist. to Center	Area	Perim.	Compact.	Mean Int.	MedR	MedG	MedB	MedH	MedS	MedV	Gray Var.	Neigh. Col. Diff.	Fore. Col. Diff.	Black Raio	Is Bckgrnd
349	225	55	0,076	29	0	218	0	0	254	1	1,000	0,996	0,0	0,380	1
229	428	80	0,067	17	246	180	16	16	15	0	0,135	0,063	48,9	0,380	0
258	176	53	0,063	30	0	218	0	0	254	1	1,000	0,996	32,1	0,380	1
250	70	30	0,077	17	178	187	13	12	9	0	0,355	0,051	276,4	0,380	0
254	125	84	0,018	84	106	41	96	98	142	0	0,040	0,569	2.748,5	0,380	1

2.3 Modeling and Training Process

In this study, the customized artificial neural network (ANN) developed using the defined superpixel-based features (T table) and automatically generated background labels (y_{bg}) was compared with MATLAB's built-in U-Net model, and the method that ensured the highest accuracy and generalizability in predicting the background class was identified.

2.3.1 Data preparation and common protocol

For all models, the input data consisted of 15 defined feature vectors (d_{center}, A, P, compactness, meanIntensity, neighbor_color_diff, foreground_color_diff, medianR, medianG, medianB, medianH, medianS, medianV, gray_variance, black_ratio). The output label was the binary variable $y_{bg} \in \{0,1\}^{N_{sp}}$ (0 = foreground, 1 = background). During model training, z-score standardization was applied to the input features. The experiments were conducted under two protocols. In the single-image protocol, each image was trained and tested only with its own superpixels, while in the combined dataset protocol, the superpixels of all images were pooled together and split into training and test sets.

2.3.2 Developed ANN model

The ANN architecture was configured with 15 neurons in the input layer (equal to the number of features), a single hidden layer with 10 neurons, and 2 neurons in the output layer representing the background/foreground classes (Figure 3). A sigmoid (logsig) activation function was employed in the hidden layer, while the output layer utilized the softmax activation function. Binary cross-entropy was adopted as the loss function, and the default Levenberg–Marquardt algorithm was selected as the optimization method. The training process was set to 200 epochs, and the model's generalizability was evaluated through 5-fold cross-validation. For each fold, independent training and testing sets were used, and the mean accuracy values were reported.

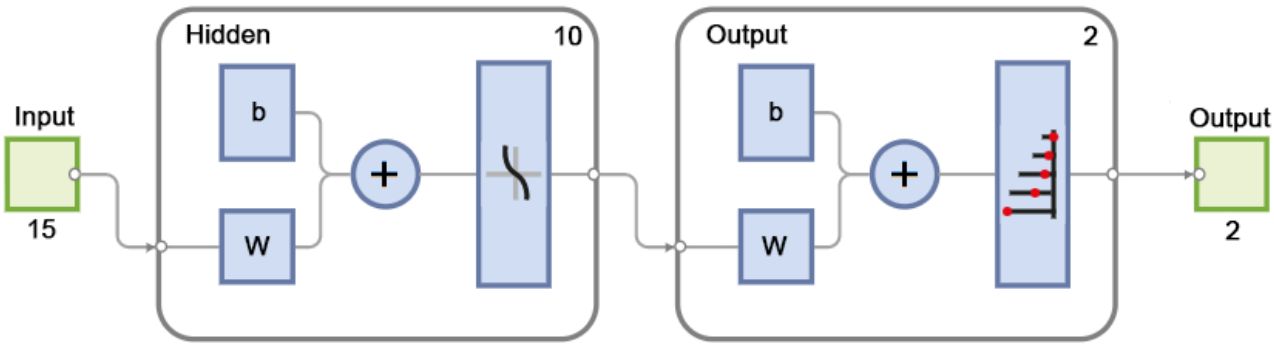

Figure 3. Architecture of the developed artificial neural network

Although the ANN and U-Net operate on different representations, superpixel-level feature vectors versus raw RGB images, both are trained on the same stripping/background annotations and evaluated on the same standardized images. Therefore, their results can be meaningfully compared in terms of practical detection performance on the common test set. This dual setup allows us to contrast a compact, feature-engineered ANN with a deep segmentation network in the same application context, highlighting the trade-offs between model complexity, data requirements, and predictive accuracy.

2.3.3 U-Net based pixel classification model

U-Net was employed for pixel-level background/foreground classification on standardized RGB inputs. Class labels were derived from annotated reference images: the 'blue mask' $1_{blue}(x, y)$ was defined using the threshold $R < 80 \wedge G < 80 \wedge B > 150$; pixels within this mask were assigned the background label, while the remaining pixels were labeled as aggregate.

Data partitioning was dynamically adjusted according to the total number of images n: if $n = 1$, the entire dataset was used for training with no validation; if $n > 1$, the dataset was randomly split into approximately 80/20 training/validation subsets. Training was conducted using the Adam optimizer with the following settings: MaxEpochs = 25, MiniBatchSize = 4, Shuffle = every-epoch; in cases where validation was available, ValidationData and ValidationFrequency = 30 were specified. No alternative techniques, such as learning rate scheduling, data augmentation, or SGDM, were applied in this study; instead, Adam was preferred for its rapid convergence in small-scale datasets.

2.4. Importance Analysis

To identify which features the model is most sensitive to and which features contribute most to its performance, a feature importance analysis was conducted. The analysis was performed exclusively on the ANN model, and each of the $p = 15$ input features was evaluated individually. The permutation importance method was employed for this analysis. In this approach, after establishing the baseline accuracy $Acc_{baseline}$, the test set values of each feature f_i were randomly permuted. The permutation disrupts the relationship between the feature and the target variable, thereby revealing the contribution of that feature to the model's predictive capability. Following each permutation, the model was re-evaluated, and the corresponding accuracy $Acc_{perm,i}$ was recorded. The importance score of each feature was then calculated as follows:

$$Importance(f_i) = Acc_{baseline} - Acc_{perm,i} \tag{12}$$

The higher this value, the greater the contribution of the corresponding feature to the model's performance.

3. RESULTS AND DISCUSSION

3.1. Findings of the Single-Image Protocol

In the single-image protocol, nine different images were employed to evaluate model performance (low level: I-5, I-18, I-25; medium level: O-2, O-22, O-28; poor level: K-3, K-8, K-9). These images were selected to represent "high," "medium," and "low" stripping percentages, with three samples included in each group. In addition, certain images deliberately contained extra objects that could be classified as noise and complicate the classification process. This design enabled the

assessment of model robustness under challenging scenarios. Figure 4 presents the Artificial Neural Network (ANN) prediction and the corresponding accuracy maps for the I-5 sample. The figure displays, in separate panels, the original image, the manually labeled ground truth mask, the ANN prediction, and the error map where misclassifications are highlighted in red.

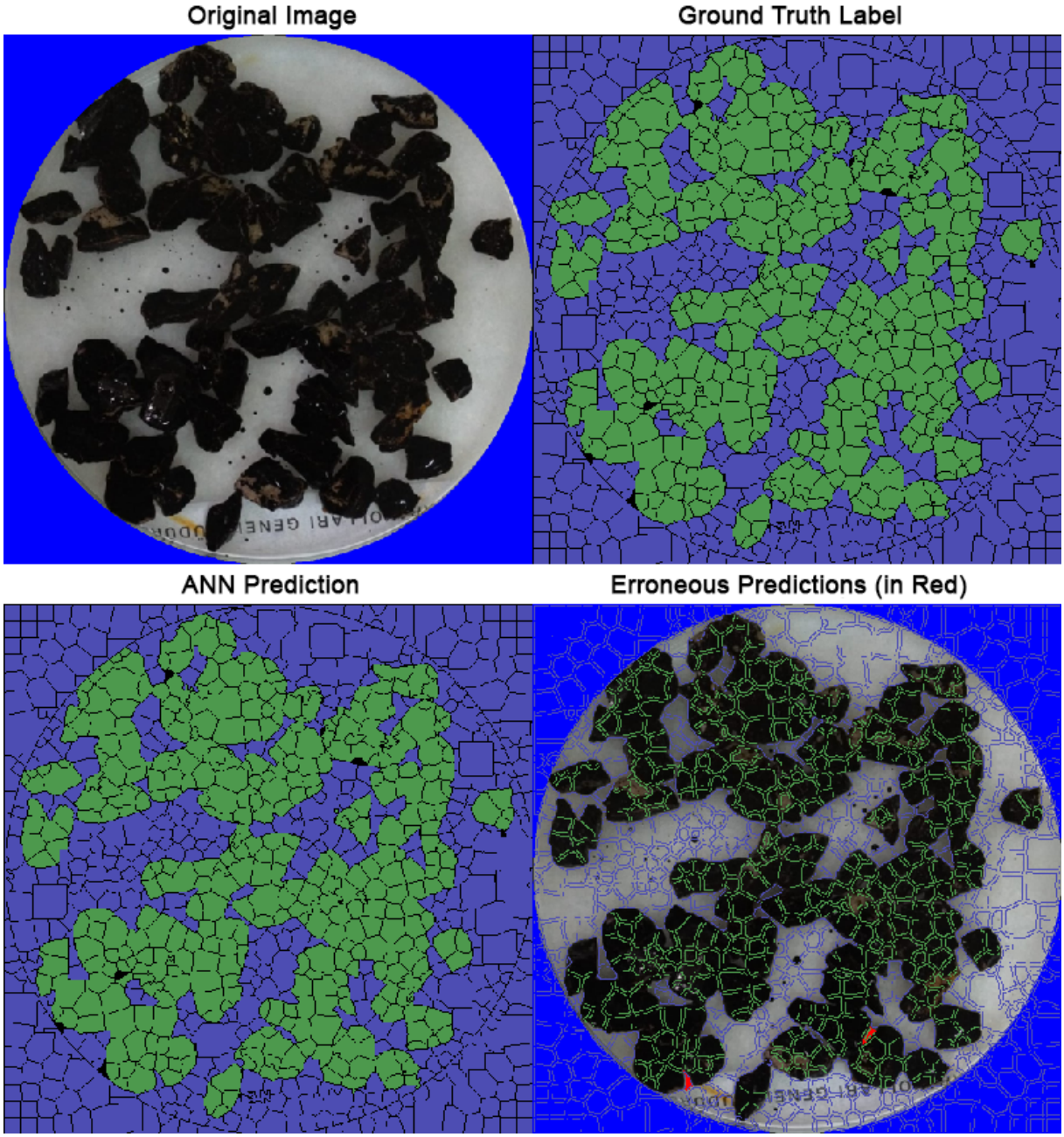

Figure 4. Results of the ANN model for the I-5 specimen; the panels show the Original Image, the Ground Truth Label (Green: Aggregate, Purple: Background), the ANN Prediction (Green Superpixel: Aggregate, Purple Superpixel: Background), and the Erroneous Predictions (in Red); each panel corresponds to a 512 × 512 pixel crop covering approximately 18 × 18 cm of the specimen surface

Similarly, Figure 5 presents the predictions of the U-Net model on the same specimen. This figure includes the original image, the ground-truth mask, the U-Net prediction, and the error map, in which misclassified pixels are marked in red.

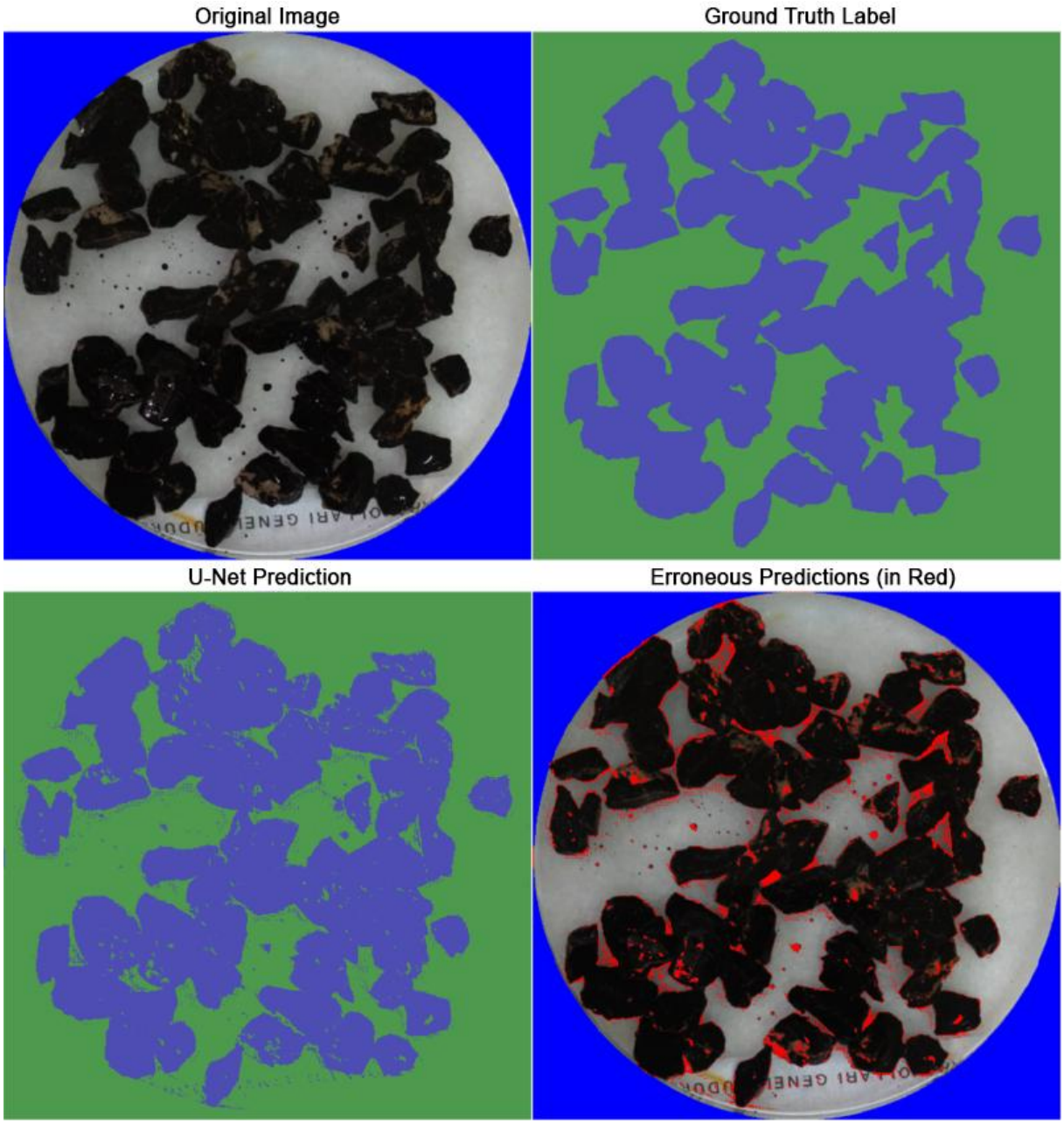

Figure 5. Results of the U-Net model for the I-5 specimen; the panels show the Original Image, the Ground Truth Label (Green: Background, Purple: Aggregate), the U-Net Prediction (Green Pixel: Background, Purple Pixel: Aggregate), and the Erroneous Predictions (in Red); each panel corresponds to a 512 × 512 pixel crop covering approximately 18 × 18 cm of the specimen surface

As observed in Figure 4, the developed ANN model did not label the small droplets inside the Petri dish, defined as noise, as aggregates. Similarly, the black text within the Petri dish was not classified as aggregate. This indicates that the model focused exclusively on the actual aggregate regions and successfully distinguished objects with similar colors but not belonging to the class. In addition, the majority of the stripping regions were correctly labeled as background, with only two small areas misclassified.

In contrast, in the U-Net predictions presented in Figure 5, droplets defined as noise and the text on the Petri dish were observed to be labeled as aggregates. Furthermore, a substantial portion of the stripping regions was classified as background, leading to deficiencies in detecting the actual stripped areas. These visual results clearly demonstrate that the developed ANN model is more robust

against noise and exhibits higher accuracy in distinguishing between bitumen and stripped regions. Such qualitative observations are also supported by the quantitative results. For the I-5 specimen, the ANN model achieved an accuracy of 99.55%, whereas the U-Net model yielded a mean Intersection over Union (Mean IoU) of 93.19% for the same data. The findings indicate that, particularly for images with complex noise, the developed ANN model provides more accurate and consistent classification compared to U-Net.

Table 2 presents the accuracy values obtained for nine specimens selected under the single-image protocol. The specimens were chosen to represent high, medium, and low levels of stripping percentage, and in some cases to include noise effects (e.g., text or droplets on the Petri dish). The same analysis procedure was applied to all images, and the results were comparatively evaluated.

Table 2. Accuracy (%) comparison of the developed ANN and U-Net models for nine different samples used in the single-image protocol

Image	Stripping Level	ANN	U-Net
I-5	Low	99.55	93.19
I-18	Low	98.10	87.91
I-25	Low	96.57	87.97
O-2	Moderate	97.79	90.27
O-22	Moderate	95.07	86.05
O-28	Moderate	91.37	78.11
K-3	High	96.48	87.75
K-8	High	97.70	91.85
K-9	High	97.38	86.56

The findings clearly demonstrate that the developed ANN model consistently outperforms the U-Net model across all samples. In particular, for samples containing noise and complex textures, such as I-5, I-18, and O-28, the accuracy difference between ANN and U-Net ranged from 6% to 13%. The highest accuracy was achieved with the ANN model for sample I-5 (99.55%), while the lowest accuracy was observed with the U-Net model for sample O-28 (78.11%). Examination of the average accuracy values further indicates that the ANN model exhibits more consistent overall performance, whereas the U-Net model shows significant deviations for certain samples. This suggests that the ANN model provides a more robust and stable learning capability against noise, particularly in the single-image protocol where only a limited number of images are available. As illustrated in Figure 6, the ANN model is capable of achieving high accuracy even on images of varying difficulty levels, with examples representing low, moderate, and high levels of peeling. Similarly, Figure 7 presents the performance of the U-Net model on the same samples, revealing that while the model performs acceptably on images with low noise, it produces unstable results when faced with noise and contrast distortions.

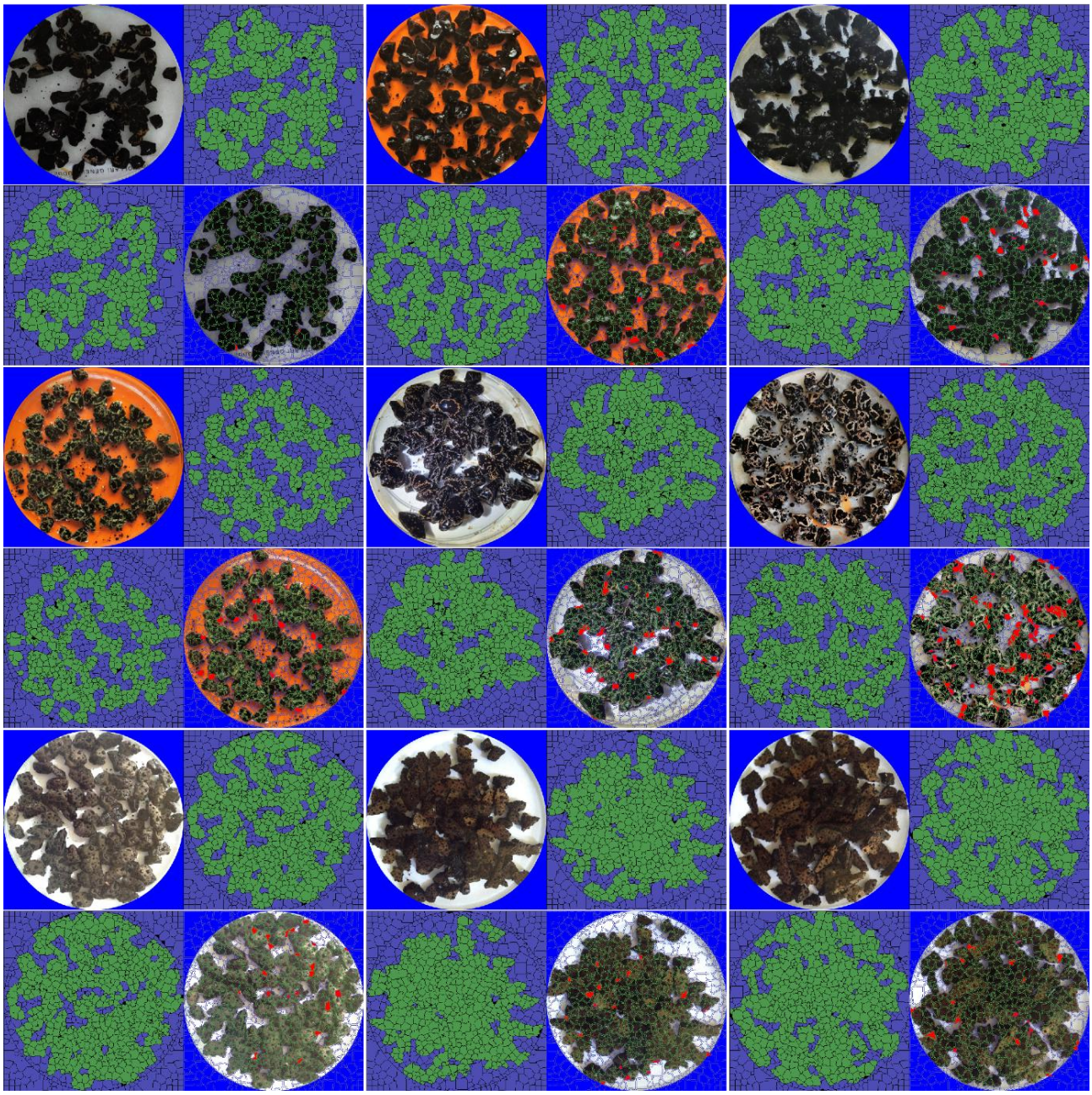

Figure 6. Prediction outputs of the ANN model across the nine specimens used in the single-image protocol: each panel shows the ANN segmentation for one specimen (green superpixels: aggregate, purple superpixels: background); each panel corresponds to a 512 × 512 pixel crop covering approximately 18 × 18 cm of the specimen surface

The visuals in Figure 6 replicate the format shown in Figure 4 (original image, labeled reference, model prediction, and error map) for each specimen. The top row contains examples representing low stripping levels, the middle row moderate levels, and the bottom row high stripping levels. When Figure 6 and Table 2 are jointly considered, it is evident that the ANN model achieved highly accurate predictions for specimens with low stripping percentages in the top row; for instance, an accuracy of 99.55% was obtained for the I-5 image. In the middle row, accuracy values declined relatively, yet were still preserved up to 91.37%. A common feature of specimens in this group is the presence of droplets generating significant noise on the Petri dish surface or stains distorting contrast. Despite these noise factors, the ANN model largely labeled bitumen-coated and stripped areas correctly, though some misclassifications occurred. Visual inspections revealed that the model did not mistakenly label non-bitumen black elements (e.g., inscriptions on the Petri dish) as bitumen,

thereby demonstrating more consistent behavior compared to the U-Net model. In the bottom row, although the noise was less prominent, the low color and texture contrast led to errors concentrated in certain small clusters. Overall, the examples in Figure 6 clearly highlight the robustness of the ANN model against noise and its ability to maintain high accuracy even across images of varying quality.

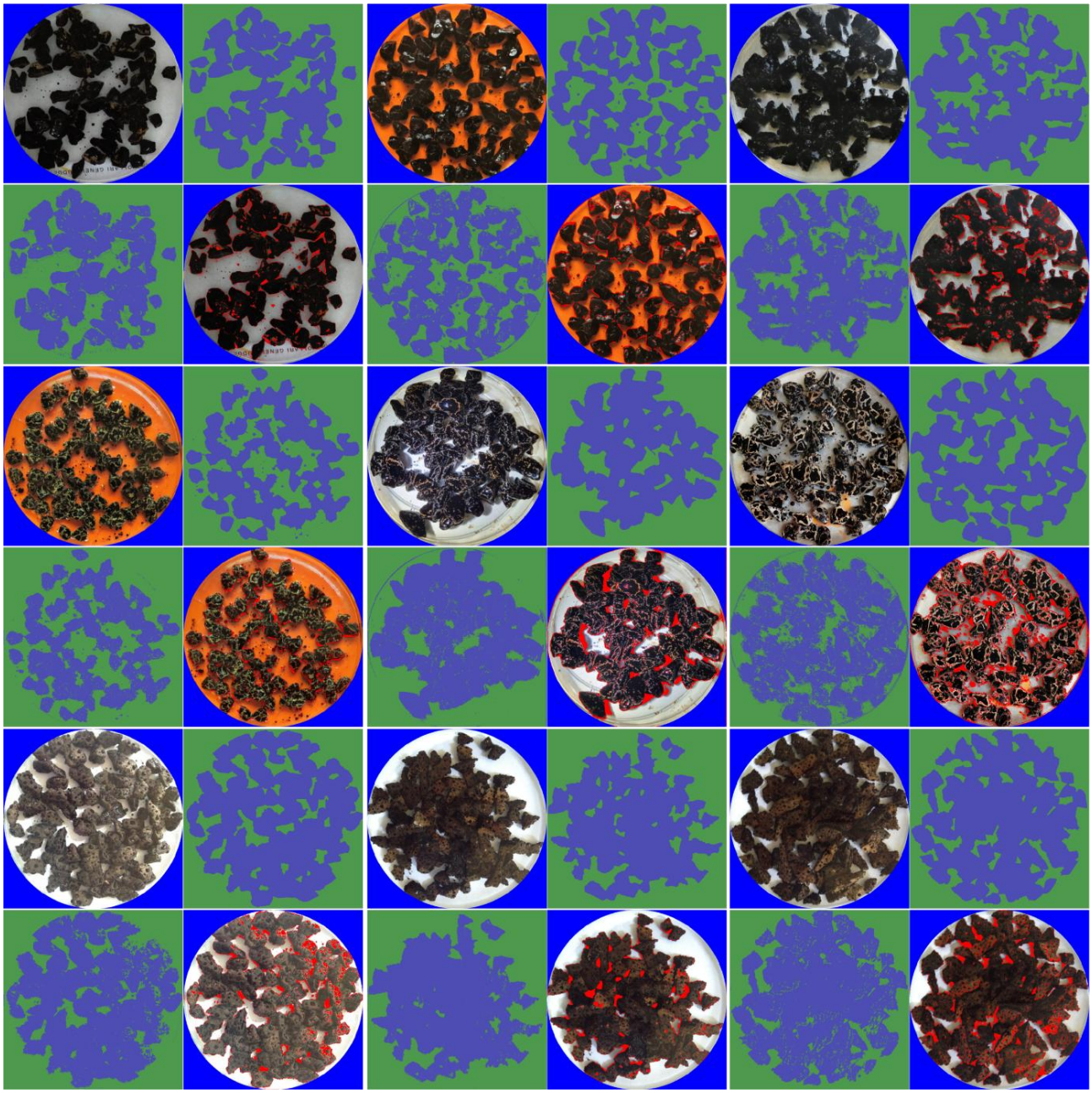

Figure 7. Prediction outputs of the U-Net model across the nine specimens used in the single-image protocol: each panel shows the U-Net segmentation for one specimen (green: background, purple: aggregate); each panel corresponds to a 512 × 512 pixel crop covering approximately 18 × 18 cm of the specimen surface

The visuals in Figure 7 were generated by applying the format presented in Figure 5 (original image, labeled reference, model prediction, and error regions) to each sample. The top row shows examples with low stripping, the middle row with moderate stripping, and the bottom row with high stripping levels. In the top row, the model generally labeled bitumen-covered areas correctly, although small errors were observed at some edges. In the middle row, which includes samples with

moderate stripping, accuracy loss is more pronounced; particularly in regions with droplets or strong reflections, the model showed uncertainty in distinguishing between bitumen and stripped areas. In the bottom row with high stripping levels, acceptable accuracy was maintained in overall shape detection, but misclassifications marked in red appeared more frequently and in clustered forms. This indicates that the U-Net model is more sensitive to noise and complex textures, and that its error rate increases under conditions of low color and texture contrast. Overall, the figure set demonstrates that the U-Net model can produce successful results under certain conditions, but its performance becomes unstable in the presence of noise and contrast distortions.

3.2. Findings of Multi-Protocol Comparison

In this section, the classification performance of the ANN and U-Net models on multi-protocol data is compared. For both models, mean accuracies and 95% confidence intervals over four independent repetitions were computed; the results are summarized in Table 3.

Table 3. Comparison of ANN and U-Net accuracies in the multi-protocol setting (mean ± 95% confidence interval over four runs)

Model	Global Accuracy (%)	Aggregate Accuracy %	Background Accuracy %
ANN	95.65 ± 0.57	97.58 ± 0.08	93.45 ± 0.67
U-Net	84.77 ± 7.79	80.87 ± 17.27	89.67 ± 4.51

As shown in Table 3, the ANN model exhibits a clear advantage under multi-protocol conditions, achieving a mean global accuracy of 95.65% (±0.57) compared to 84.77% (±7.79) for U-Net. In the class-based evaluation, ANN attains a notably high accuracy of 97.58% (±0.08) for the aggregate class, whereas U-Net remains at 80.87% (±17.27) in the same category. For the background class, both models reach high accuracy levels, with ANN (93.45% ± 0.67) again slightly outperforming U-Net (89.67% ± 4.51). The substantially narrower confidence intervals of the ANN model, particularly for the aggregate class, indicate a much more stable and reproducible performance across independent runs, while the wide intervals observed for U-Net reveal a strong sensitivity to initialization and data splits. These results indicate that ANN provides a more balanced and stable performance across classes, particularly yielding more reliable predictions in distinguishing between bitumen-coated and stripped areas. The distribution of class predictions, error rates, and overall performance metrics of the ANN model are further examined through the confusion matrices in Figure 8, while the evolution of accuracy and loss during U-Net training is presented in Figure 9.

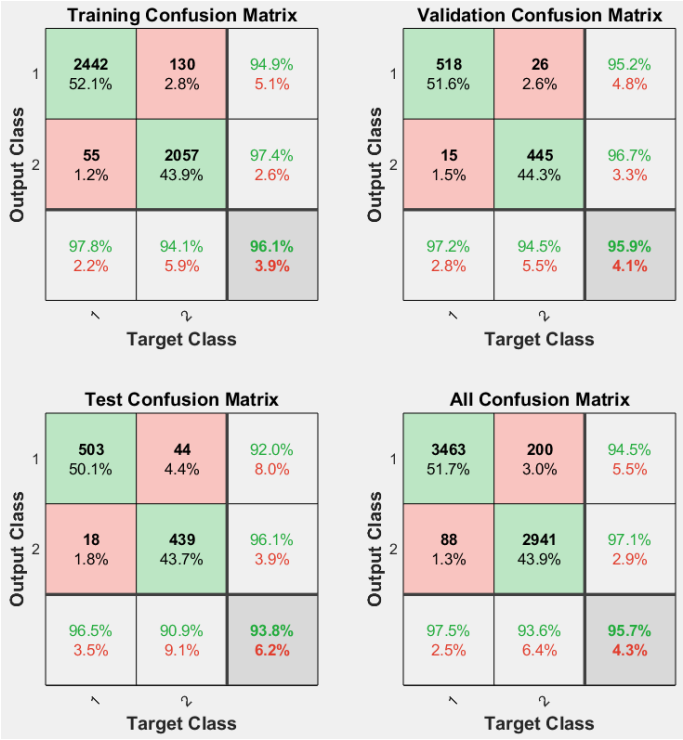

Figure 8. Confusion matrices of the ANN model obtained from training, validation, test, and overall data under the multi-protocol setting

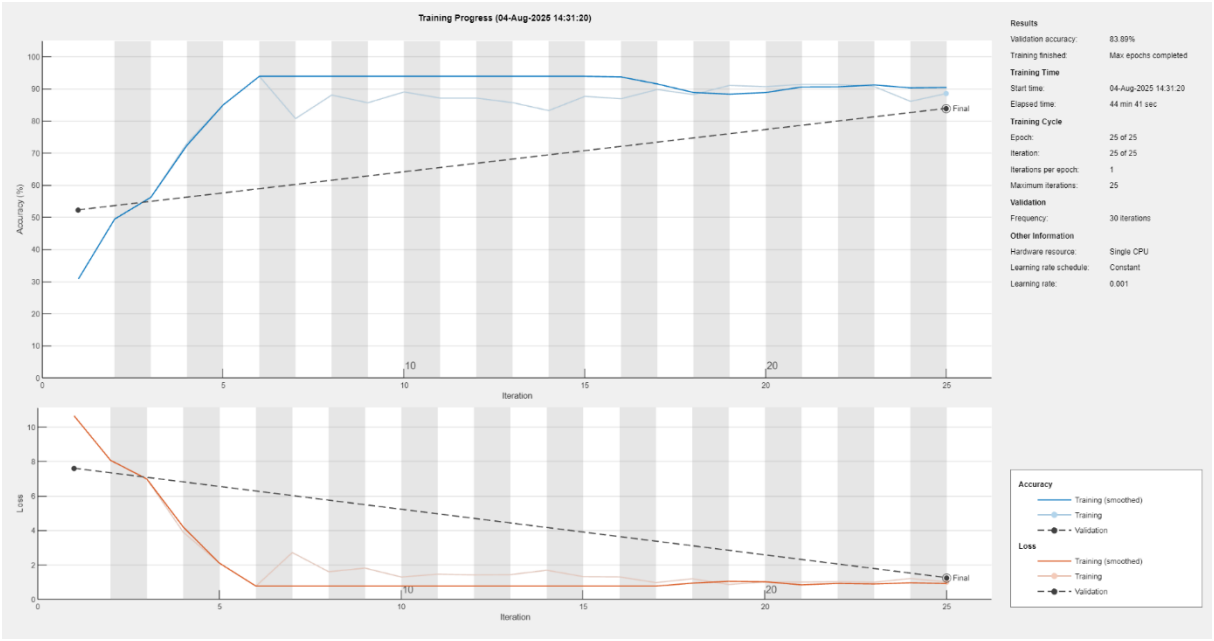

Figure 9. Training and validation loss/accuracy curves of the U-Net model under the multi-protocol setting

As observed in Figure 8, the error distribution matrix of the ANN indicates that misclassifications were limited both in quantity and variety, suggesting a strong generalization capability of the model. In contrast, Figure 9 shows that the error curves of U-Net reveal a plateau in accuracy gains after a certain stage of training, indicating constrained improvement. These findings highlight that, within multi-protocol scenarios, ANN appears to provide a more reliable option.

To complement the accuracy-based comparison, we also report additional threshold-based metrics for the ANN in the multi-protocol setting. Aggregating the confusion-matrix terms over the five cross-validation folds and treating the background class as the positive class, the ANN achieved

an overall accuracy of 95.65%, a precision of 0.961, a recall (sensitivity) of 0.943, a specificity of 0.966, an F1-score of 0.952, and an IoU of 0.908 for the positive class. These values indicate a well-balanced trade-off between false positives and false negatives and further confirm the robustness of the ANN model in separating background from aggregate regions.

3.3. Importance Analysis

The relative importance levels of the features influencing the model's classification performance are presented in Figure 10. According to the analysis results, medianV (the brightness component in the HSV color space) has the highest importance score, with a mean value of 0.19 ± 0.01 over four independent runs (visually shown as approximately 0.20 in Figure 10). This confirms its decisive role in the model's decision-making process. It is followed by medianB (the median of the blue channel in the RGB color space) and medianS (saturation), indicating that color- and brightness-based features directly contribute to classification accuracy.

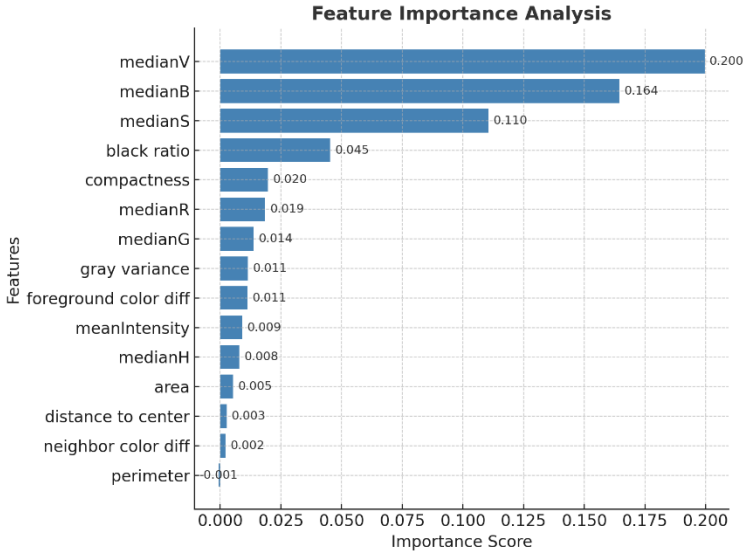

Figure 10. Relative importance levels of the features

On the other hand, the black_ratio variable also provides a meaningful contribution (~0.045). This indicates that the proportion of black color in objects is an important discriminative factor, particularly in detecting bituminous surfaces. In contrast, geometry-based features such as distance_center, area, and perimeter exhibit considerably low importance scores.

These findings reveal that the model predominantly relies on color- and brightness-oriented decisions, while shape and spatial features make only a limited contribution. Therefore, to further enhance classification performance, it would be an appropriate strategy to focus on improving color-based features and, if necessary, to consider additional spectral transformations.

The findings regarding feature utilization also demonstrate clear distinctions between different model types such as ANN and U-Net. Since the ANN model is fed directly with numerical feature vectors, the contribution of each feature to the classification decision can be explicitly computed and interpreted. In contrast, the U-Net model takes raw image pixels as input and, through its multi-layered convolutional structure, automatically learns color, texture, and edge information within its intermediate layers. Therefore, unlike in ANN, the relative importance of predefined features in U-Net cannot be directly obtained; instead, techniques such as Grad-CAM or similar visualization methods are employed to analyze which visual regions the model attends to. This distinction arises

from the higher interpretability of ANN and the broader learning capacity of U-Net, though it precludes direct feature-level comparison between the two approaches.

It is important to address the methodological differences in the comparison. Although U-Net is a more expressive architecture with a higher representational capacity, its training in this study was intentionally kept simple: a standard encoder–decoder configuration with Adam optimization, fixed learning rate, and no data augmentation or learning-rate scheduling. Under these small-data conditions (nine standardized RGB images) and without aggressive regularization, the pixel-wise U-Net model proved more sensitive to noise and exhibited larger variability in accuracy, whereas the feature-based ANN remained markedly more data-efficient and stable. Therefore, the lower performance of U-Net in this work should be interpreted as a limitation of the current training regime and dataset size, rather than as a fundamental weakness of the architecture for stripping detection problems. This comparison highlights that for engineering applications with limited labeled data, feature-engineering approaches (like the proposed superpixel-ANN) can offer a more robust 'ready-to-use' solution than deep learning models requiring extensive optimization. Investigating more advanced optimization strategies and larger, more diverse datasets for U-Net constitutes a natural extension of the present study.

4. CONCLUSIONS

Within the scope of this study, the developed model successfully distinguished aggregate and background in stripping tests applied during asphalt mixture production through visual analysis. For this purpose, the images of the test specimens were standardized, manually labeled, and thereby used to construct training datasets. The classification tasks were carried out separately using a simple Artificial Neural Network (ANN)-based model and the U-Net deep learning architecture. The ANN model was designed as a single-layer structure with 10 neurons, and different normalization techniques (z-score and min-max) were also tested for data scaling.

The ANN model achieved higher accuracy compared to the U-Net model in both single-image and multi-image protocols. In particular, for the single-image analysis of specimen I-5, the ANN model reached an accuracy of 99.55%, whereas the U-Net model remained at 93.19%. Likewise, under the multi-image protocol, the average accuracy calculated across all specimens was approximately 95.7% for the ANN model, while it was around 83.89% for U-Net. These results demonstrate that the proposed simple ANN-based approach exhibits a clear performance advantage, even compared to a complex deep network architecture such as U-Net, which is widely employed for pixel-based segmentation.

The analysis of the results revealed that the superior performance of the ANN model (Global Accuracy 95.7%) compared to U-Net (83.89%) stems primarily from the robustness of the superpixel-based feature extraction process and the high discriminative power of the selected features. The Feature Importance Analysis (Section 3.3) showed that the ANN model largely relied on color- and brightness-based features (medianV, medianB, medianS) in the decision-making process. The distinct differences in color and brightness between the dark bituminous aggregate material and the light/blue background surface explain why such a separation could be achieved with high accuracy even by a relatively simple ANN structure. The superpixel approach aggregates noisy pixel data into stable, statistically representative units, rendering the ANN model more stable against noise and contrast distortions than the pixel-based U-Net, particularly as confirmed by the Single-Image Protocol findings (Figure 6 vs. Figure 7). Furthermore, the model’s design utilizes contextual descriptors, such

as the black pixel ratio, which enhances robustness by explicitly incorporating scene-specific information often missed by raw pixel input models like U-Net. The finding that alternative architectures or different normalization techniques did not yield meaningful improvement suggests that, owing to the high discriminative power of the problem-specific features, even a simple single-layer network architecture is sufficient, and more complex ANN structures are not currently necessary. A limitation of our study is that the ANN and U-Net models rely on different input representations (hand-crafted superpixel features versus raw image data). As a result, their comparison should be interpreted as a practical assessment of two alternative modeling paradigms for stripping detection, rather than as a strictly controlled architectural benchmark.

Overall, the proposed method provides a highly accurate and practically flexible solution. Due to the low computational complexity of the superpixel-based feature extraction, the developed ANN model has been successfully integrated into a standalone software application that operates efficiently on standard laboratory computers without the need for high-performance GPUs. Furthermore, the geometric standardization step ensures that images captured by ubiquitous devices, such as mobile phone cameras, can be consistently processed. This makes the method a readily deployable tool for routine stripping analysis in field or laboratory settings, independent of expensive imaging setups. While the number of raw images used (nine samples) is limited, the superpixel-based feature extraction process, which generates approximately 9000 independent feature vectors, ensures sufficient data points for robust model training and prevents overfitting. However, it is considered that incorporating certain preprocessing steps or data augmentation techniques could further enhance classification performance and address generalization concerns for extremely diverse, large-scale datasets. In particular, removing noise or artifacts from the images may enable the model to perceive discriminative features more clearly, thereby contributing positively to accuracy. Indeed, the literature reports that noise reduction techniques applied in fields such as medical image analysis have increased the accuracy of deep learning-based models by approximately 3.5% (Mechria, Hassine, & Gouider, 2022). The findings of the present study are consistent with this evidence, and similarly, it is anticipated that incorporating additional preprocessing steps such as noise reduction, or implementing data augmentation, which was not applied in this study, could further strengthen our model for aggregate–background separation. In conclusion, this study, which provides an automatic visual analysis as an alternative to expert evaluation, supports comparable approaches in the literature with its high accuracy and implementation flexibility, and offers promising potential for future research.

5. CONFLICT OF INTEREST

Authors approve that to the best of their knowledge, there is not any conflict of interest or common interest with an institution/organization or a person that may affect the review process of the paper.

6. AUTHOR CONTRIBUTION

Kadir AKGÖL and Mehmet Can TUNA contributed to the determining and management concept and/or design process of the reseach, data collection, data analysis and interpretation of the results, preparation of the manuscript, critical analysis of the intellectual content, final approval and full responsibility.

7. REFERENCES

Cao R. J., Zhao Y. L., Gao Y., Huang X. M., Zhang L. L., Effects of flow rates and layer thicknesses for aggregate conveying process on the prediction accuracy of aggregate gradation by image segmentation based on machine vision. Construction and Building Materials 222, 566-578, 2019.

Cui P. D., Wu S. P., Xiao Y., Wang F., Wang F. S., Quantitative evaluation of active based adhesion in Aggregate-Asphalt by digital image analysis. Journal of Adhesion Science and Technology 33(14), 1544-1557, 2019.

Gürer C., Karaşahin M., Sathi Kaplama Agregalarının Adezyon Özelliklerinin Araştırılması. Yapı Teknolojileri Elektronik Dergisi 10(2), 1-11, 2016.

Huang H. H., Luo J. Y., Tutumluer E., Hart J. M., Stolba A. J., Automated Segmentation and Morphological Analyses of Stockpile Aggregate Images using Deep Convolutional Neural Networks. Transportation Research Record 2674(10), 285-298, 2020.

Huang T., Liu G. Q., Evaluation for coarse aggregate distribution of asphalt mixtures based on the two-dimensional digital image analysis. Construction and Building Materials 450, 138716, 2024.

Kamani M., Ajalloeian R., Investigation of the changes in aggregate morphology during different aggregate abrasion/degradation tests using image analysis. Construction and Building Materials 314, 125614, 2022.

Li H. J., Asbjörnsson G., Lindqvist M., Image Process of Rock Size Distribution Using DexiNed-Based Neural Network. Minerals 11(7), 736, 2021.

Li M., Wang J., Guo Z. B., Chen J. C., Zhao Z. D., Ren J. L., Evaluation of the Adhesion between Aggregate and Asphalt Binder Based on Image Processing Techniques Considering Aggregate Characteristics. Materials 16(14), 5097, 2023.

Mechria H., Hassine K., Gouider M. S., Effect of Denoising on Performance of Deep Convolutional Neural Network for Mammogram Images Classification. Procedia Computer Science 207, 2345-2352, 2022.

Öner J., Seramik Atıklarıyla Hazırlanan Asfalt Karışımların Soyulmaya Karşı Dayanımının Belirlenmesi. Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 20(3), 498-505, 2020.

Peng Y. J., Ying L. P., Kamel M. M. A., Wang Y., Mesoscale fracture analysis of recycled aggregate concrete based on digital image processing technique. Structural Concrete, 22, E33-E47, 2020.

Reddy G. S., Abdallah I. N., Nazarian S., Contributions of aggregate mineralogical and morphological parameters to aggregate frictional performance. Construction and Building Materials 478, 141413, 2025.

Reyes-Ortiz O. J., Mejia M., Useche-Castelblanco J. S., Digital image analysis applied in asphalt mixtures for sieve size curve reconstruction and aggregate distribution homogeneity. International Journal of Pavement Research and Technology 14(3), 288-298, 2021.

Salemi M., Wang H., Image-aided random aggregate packing for computational modeling of asphalt concrete microstructure. Construction and Building Materials 177, 467-476, 2018.

Sinecen M., Makinaci M., Classification of Aggregates Using Basic Shape Parameters Through Neural Networks. Pamukkale University Journal of Engineering Sciences-Pamukkale Universitesi Muhendislik Bilimleri Dergisi 16(2), 149-153, 2010.

Sinecen M., Makinaci M., Topal A., Aggregate Classification by Using 3D Image Analysis Technique. Gazi University Journal of Science 24(4), 773-780, 2011.

Théodon L., Coufort-Saudejaud C., Hamieh A., Debayle J., Morphological characterization of compact aggregates using image analysis and a geometrical stochastic 3D model, Paper presented at the IEEE 13th International Conference on Pattern Recognition Systems (ICPRS), Guayaquil, ECUADOR, July 04-07, 2023.

Wang H. N., Wang C. H., Bu Y., You Z. P., Yang X., Oeser M., Correlate aggregate angularity characteristics to the skid resistance of asphalt pavement based on image analysis technology. Construction and Building Materials 242, 118150, 2020.

Wang L. B., Lane D. S., Lu Y., Druta C., Portable Image Analysis System for Characterizing Aggregate Morphology. Transportation Research Record 2104(1), 3-11, 2009.

Xiao R., Polaczyk P., Huang B. S., Measuring moisture damage of asphalt mixtures: The development of a new modified boiling test based on color image processing. Measurement 190, 110699, 2022.

Xing C., Xu H. N., Tan Y. Q., Liu X. Y., Ye Q., Mesostructured property of aggregate disruption in asphalt mixture based on digital image processing method. Construction and Building Materials 200, 781-789, 2019.

Yan R., Liao J. D., Wu X. Y., Xie C. J., Xia L., Research on Classification Method of Sand and Gravel Aggregate Based on Convolutional Neural Network. Laser & Optoelectronics Progress 58(20), 2021.

Zong S. L., Zhou G. Z., Li M., Wang X. Z., Deep learning-based on-line image analysis for continuous industrial crystallization processes. Particuology 74, 173-183, 2023.