Karadeniz Fen Bilimleri Dergisi, 15(4), 1647-1667, 2025. DOI: 10.31466/kfbd.1704724

KFBD

Karadeniz Fen Bilimleri Dergisi
The Black Sea Journal of Sciences
ISSN (Online): 2564-7377 https://dergipark.org.tr/tr/pub/kfbd

Araştırma Makalesi / Research Article

Evaluation of Stripping Test Outcomes Using Artificial Neural Networks: A Comparative Analysis of Image Processing and Expert Assessments

Kadir AKGOL[1], Ferdi ÖZBİLGİN[2], Omer Faruk OZTURK[3*], Mehmet Can TUNA[4], Mücahit KAYİŞ[5], Fatih ULU[6]

Abstract

In this study, the stripping behavior of aggregates used in flexible pavement layers was analyzed by comparing laboratory test results with image processing techniques and expert evaluations. A total of 270 aggregate images obtained under laboratory conditions were analyzed using a custom tool developed in MATLAB. This tool extracted the total stripped area, stripped perimeter, and number of stripped regions for each image and subsequently calculated the stripping percentage. The same specimens were also visually evaluated by laboratory personnel, and expert-based assessments were recorded. The accuracy and consistency of the obtained data were tested using the Artificial Neural Network (ANN) method. The results of this study demonstrate that image processing techniques offer a more objective and reproducible alternative for evaluating stripping behavior. Moreover, the findings highlight the necessity of considering the geometric characteristics of stripped regions in addition to the total stripped area. These insights serve as a guide for improving the accuracy of laboratory stripping tests.

Keywords: Image process, Artificial neural networks, Stripping test, Aggregate-Bitumen adhesion.

Soyulma Deneyi Sonuçlarının Yapay Sinir Ağları ile Değerlendirilmesi: Görüntü İşleme ve Uzman Görüşü Karşılaştırması

Öz

Bu çalışmada, karayolu üst yapılarında kullanılan agregaların soyulma davranışını analiz etmek amacıyla laboratuvar deney sonuçları, görüntü işleme teknikleri ve uzman değerlendirmeleri temelinde karşılaştırılmıştır. Çalışma kapsamında, laboratuvar ortamında elde edilen 270 adet agrega görüntüsü MATLAB yazılımında geliştirilen bir araç kullanılarak analiz edilmiştir. Bu araç, her bir görüntü için toplam soyulma alanı, soyulma çevresi ve soyulma miktarını belirlemiş, ardından soyulma yüzdesini hesaplamıştır. Aynı numuneler, laboratuvar personeli tarafından görsel olarak değerlendirilerek uzman görüşüne dayalı sonuçlar kaydedilmiştir. Elde edilen verilerin doğruluğu ve tutarlılığı yapay sinir ağları (YSA) yöntemi ile test edilmiştir. Bu çalışmanın sonuçları, görüntü işleme tekniklerinin soyulma davranışının nesnel ve tekrarlanabilir bir şekilde değerlendirilmesi için daha güvenilir bir alternatif sunduğunu ortaya koymaktadır. Ayrıca, toplam soyulma alanına ek olarak soyulma bölgelerinin geometrik özelliklerinin değerlendirilmesi gerektiği vurgulanmaktadır. Bu bulgular, laboratuvar deneylerinin doğruluğunu artırmak için yol gösterici olmuştur.

Anahtar Kelimeler: Görüntü işleme, Yapay sinir ağları, Soyulma deneyi, Agrega-Bitüm Adezyonu.

[1,3,4,5,6]Giresun University, Civil Engineering Department, Giresun, Türkiye, kadir.akgol@giresun.edu.tr omer.ozturk@giresun.edu.tr mehmetcantunaa@hotmail.com mucahitkayis@outlook.com flu1453@gmail.com
[2]Giresun University, Electrical-Electronic Engineering Department, Giresun, Türkiye, ferdi.ozbilgin@giresun.edu.tr

[*]Sorumlu Yazar/Corresponding Author

Geliş/Received: 26.05.2025 **Kabul/Accepted:** 26.09.2025 **Yayın/Published:** 15.12.2025

Akgol, K., Özbilgin, F., Ozturk, O.F., Tuna, M.C., Kayiş, M., Ulu, F. (2025). Evaluation of Stripping Test Outcomes Using Artificial Neural Networks: A Comparative Analysis of Image Processing and Expert Assessments. *Karadeniz Fen Bilimleri Dergisi*, 15(4), 1647-1667.

1. Introduction

The performance and durability of flexible pavement structures remain a critical subject of ongoing research due to the influence of traffic loads and environmental factors. Aggregates and bitumen used in hot mix asphalt are subjected to a series of laboratory tests before field application to evaluate their suitability and performance (White et al., 2006). One of the most critical tests employed in this evaluation process is the stripping test, which aims to measure the adhesion capacity between aggregate and bitumen (Bundy et al., 2000). The results of the stripping test play a significant role in determining the bonding performance of aggregates used in asphalt mixtures, thereby contributing to the assessment of factors affecting pavement service life (Valentova et al., 2015). These results are generally expressed as the ratio of the stripped area to the total aggregate surface area. When necessary, they also provide the basis for decisions regarding the addition of adhesion-promoting additives to the binder (Simons et al., 2005). However, since the evaluation is typically based on visual inspection, the outcomes may vary depending on the evaluator's level of expertise, experience, and subjective judgment, which may, in turn, compromise the reliability of the measurements.

One of the most critical parameters affecting the structural durability of asphalt pavements is the adhesion force between the aggregate and the bitumen. A reduction in this adhesion may lead to premature surface distresses and loss of performance. Several factors influencing adhesion strength have been examined in the literature. Gürer and Karaşahin (2014) emphasized that the structural properties of aggregates significantly affect their bonding performance with bitumen, noting that volcanic or silica-rich aggregates tend to exhibit weaker adhesion. Karadağ and Saltan (2021) investigated the potential of chemical additives such as Cocamide Diethanolamide to improve bitumen-aggregate adhesion, reporting a noticeable enhancement in adhesion with increasing additive content. In addition, the use of modified bitumen incorporating recycled materials has also been studied. Akalın and Karacasu (2020) found that modifying bitumen with environmentally friendly additives improved adhesion and enhanced stripping resistance in surface treatments. Similarly, Girdap (2017) reported that the use of "Pr Plast S" additive in Stone Mastic Asphalt mixtures improved aggregate-bitumen bonding performance. While aggregate type and texture have been evaluated from various perspectives, Sajjadi (2016) concluded that polymer-modified bitumens, such as those containing SBS, yield better adhesion performance compared to other additives. Furthermore, Öner (2020) demonstrated through microscopic analysis that the use of ceramic waste as aggregate material positively influences bitumen-aggregate bonding and stripping resistance in asphalt mixtures.

Inherent subjectivity and reproducibility issues in conventional evaluation methods have necessitated the development of novel technological approaches for analyzing the performance of asphalt pavements. In recent years, image processing techniques and AI-based methods have emerged as effective tools for the objective analysis of stripping behavior and aggregate-bitumen interactions. Oh and Barham (2011) highlighted the advantages of digital imaging technologies in pavement performance evaluation, asserting that these methods offer significantly higher accuracy and objectivity compared to traditional visual assessments. Similarly, Arasan (2011) analyzed the impact of aggregate shape on mechanical properties using digital imaging methods, demonstrating that the shape factor of aggregates directly influences the performance of asphalt concrete. In addition, recent studies employing deep learning, a data-driven AI approach, have shown that such methods yield more accurate and reliable results in detecting and analyzing surface deformations and distresses in asphalt pavements (Yang et al., 2022).

The application of image processing techniques in the analysis of pavement distresses is widely documented in the literature. Çetin (2012) classified pavement surface damages and deformations using image processing methods and developed automated analysis programs to effectively detect these distresses. In a similar vein, Aktaş (2012) performed an analysis of deformations in surface layers through image processing techniques. Doğan and Ergen (2022) applied an edge-based image processing algorithm, a cost-effective approach, to detect asphalt cracks, while Rajab et al. (2008) employed ImageJ software to quantify road surface distresses, demonstrating that digital measurements are consistent with manual assessments and establishing the method as a reliable alternative. Furthermore, Shen (2016) developed an image processing-based crack library to automate the classification of different pavement crack types, and Lee et al. (2018) utilized GPS-supported image processing techniques to detect potholes on highway surfaces, thereby demonstrating significant practical advantages. Additionally, Mo et al. (2020) developed an application to identify violations of truck weight limits using image processing techniques for highway safety, and Wigan (1992) evaluated the overall applicability of image processing in road and traffic research, illustrating that these methods can be successfully used for rating pavement conditions and automating the classification of surface defects.

Image processing techniques are widely used not only for asphalt surface analysis but also in traffic monitoring and road safety applications. One of the most prominent areas of application for these techniques is vehicle classification and traffic flow analysis. Niksaz (2012) employed image processing techniques to perform automated vehicle counting, thereby enabling real-time traffic flow monitoring and congestion prediction. In a study by Sarikan et al. (2017), different types of vehicles were automatically classified using image processing methods, and the distinction between motorcycles and sedan vehicles was accurately detected. For the rapid and accurate detection of traffic

incidents, Bubeníková et al. (2015) utilized digital video processing techniques to automatically identify highway traffic events. Furthermore, Kyo et al. (1999) developed a real-time detection system based on a dedicated image processing board specifically designed for adverse weather conditions. These studies clearly demonstrate that image processing techniques offer effective and reliable solutions for traffic and road safety management systems.

In recent years, beyond image-based methods, artificial intelligence techniques, particularly artificial neural networks, have gained prominence in the evaluation and management of structural performance in asphalt pavements. Artificial neural networks (ANNs) are data-driven analytical tools that excel in modeling complex and nonlinear relationships. Beltran and Romo (2014) reported that ANNs outperform traditional models in predicting asphalt pavement performance and maintenance planning, providing high levels of accuracy. Similarly, Sollazzo et al. (2017) conducted studies on optimizing pavement performance using neural networks, showing that data-driven approaches produce more reliable and accurate predictions than conventional methods. These findings reinforce the significance and effectiveness of AI-based methods in modern road infrastructure management systems.

A review of the existing literature reveals that both conventional methods and data-driven approaches have been employed in various applications to assess the performance of asphalt pavements. However, studies that simultaneously utilize image processing techniques and artificial neural networks to analyze aggregate–bitumen adhesion in asphalt pavements remain limited. Moreover, it is evident that in most previous studies, the evaluation of stripping tests has predominantly relied on expert opinion-based subjective methods, with limited emphasis on assessing the accuracy and consistency of these approaches. This study aims to address this gap by proposing a model that integrates image processing and artificial neural networks to conduct more objective and reliable stripping analyses for asphalt pavements. To identify the limitations of existing evaluation methods and to develop a new assessment approach, the study conducts a comparative analysis of data obtained from expert evaluations and image processing techniques. This comparison allows for the identification of consistencies and discrepancies between methods and leads to the development of a more objective evaluation model for analyzing stripping behavior.

2. Materials and Methods

This section presents the dataset construction, exploratory data analysis, and the methodology employed to develop the prediction model. The process includes the acquisition and categorization of aggregate images, statistical evaluation of input variables, and the design and implementation of

an Artificial Neural Network (ANN) model for estimating stripping rates. Each stage is detailed in the following subsections.

2.1. Data Acquisition

As part of the experimental studies, photographs of 30 aggregate specimens exhibiting varying degrees of stripping and image quality were obtained from the laboratories of the General Directorate of Highways. These images were systematically organized into specific categories and subjected to a series of preprocessing steps. In the first stage, the original 30 images were analyzed without any cropping or modification. In the second stage, each original image was segmented into four halves (top, bottom, left, and right), resulting in 120 half-images. In the third and final stage, each original image was divided into four quadrants, generating an additional 120 quarter-images. Through this multi-step segmentation process, nine derivative images were generated for each original sample, yielding a comprehensive dataset of 270 images in total (Figure 1). All images were visually evaluated by expert laboratory personnel who conducted the original stripping tests, and the corresponding stripping percentages were recorded for each image.

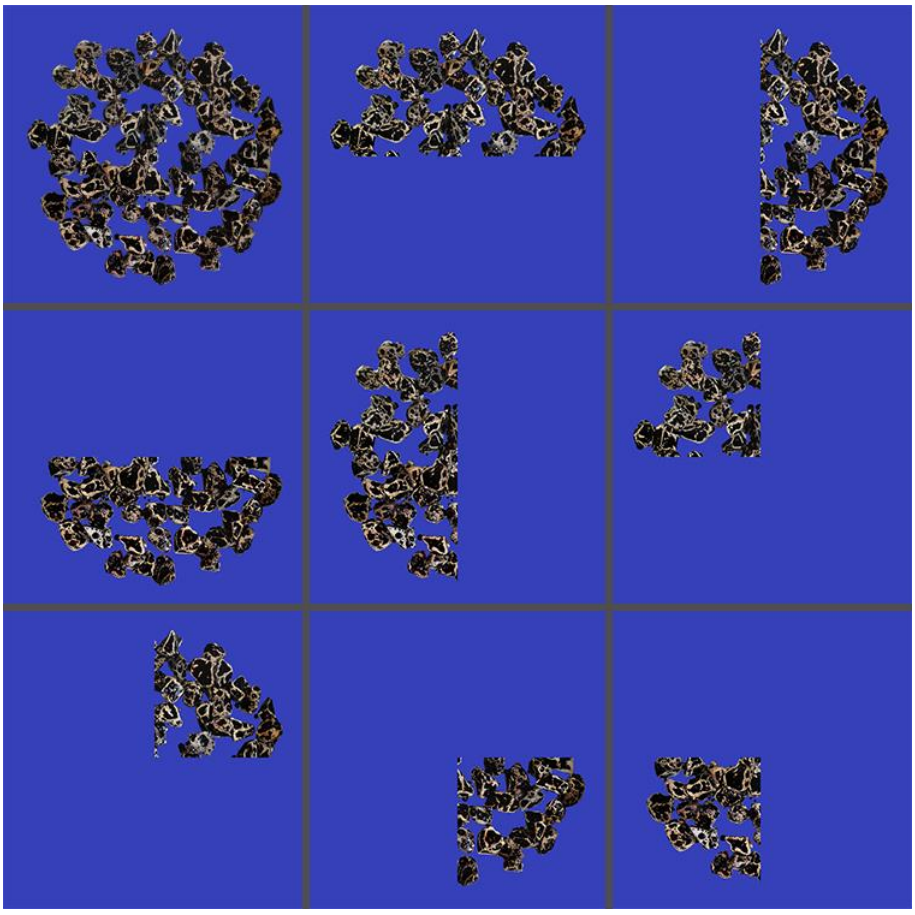

Figure 1. Sample images categorized by processing type.

To automatically determine stripping percentages on the 270 generated images, a custom software tool was developed using the MATLAB Image Processing Toolbox. This tool applies various segmentation techniques to calculate, for each specimen, the total aggregate surface area, the stripped areas on the aggregate, the perimeters of these stripped regions, and their centroids. As an example, the application of image processing techniques on a selected specimen is illustrated in Figure 2.

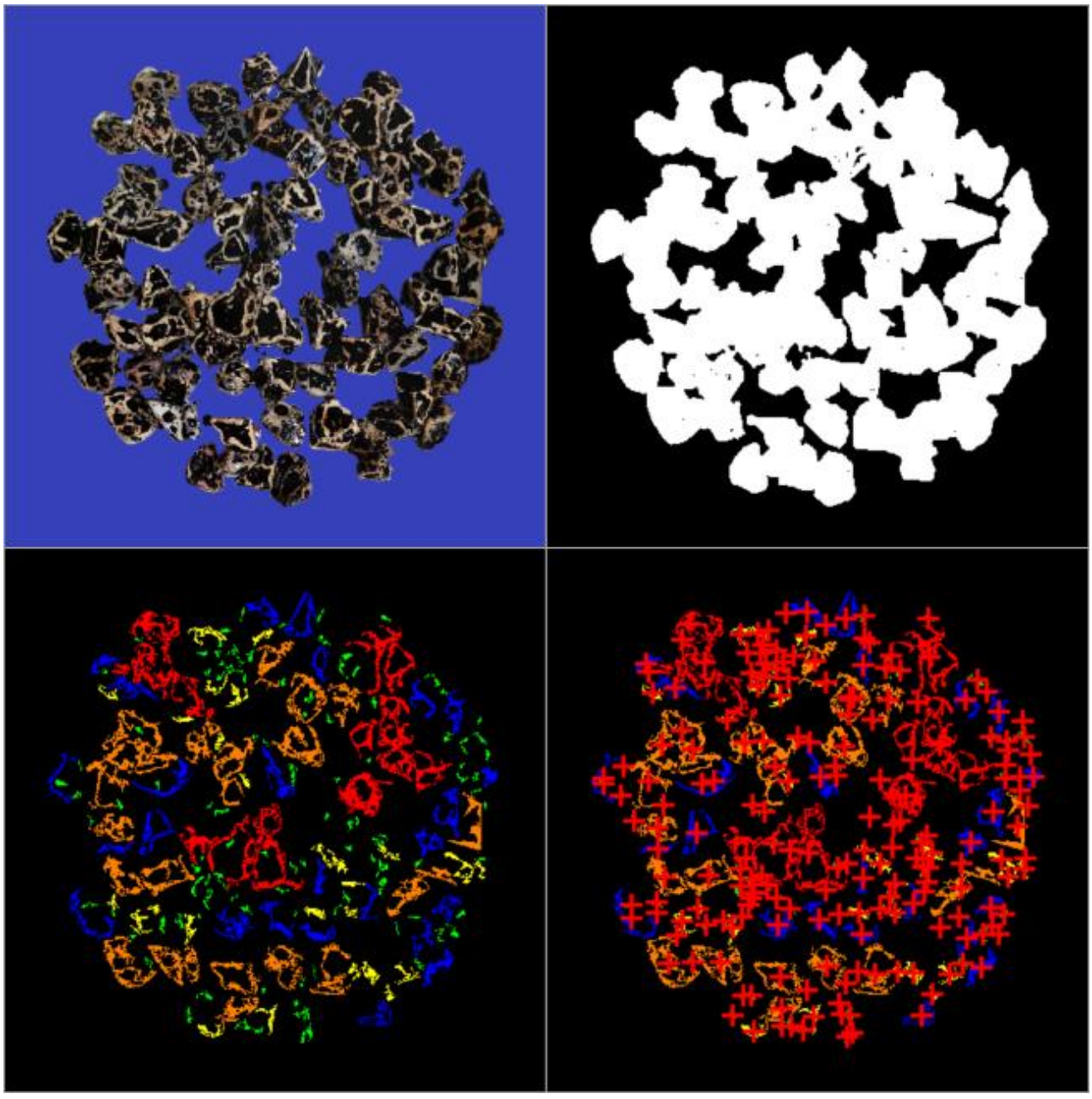

Figure 2. Example of image processing applied to an aggregate specimen.

As illustrated in Figure 2, each stripped region was individually identified, and its area was calculated in pixels and color-coded based on its size. Based on the data produced by the software,

the total stripped area, total stripped perimeter, and total number of stripped regions were calculated. The stripping percentage was then determined by calculating the ratio of the total stripped area to the aggregate surface area. This approach enabled a comparative analysis of the stripping percentages obtained through two different methods, expert evaluation and image processing, for all 270 data samples.

In recent years, artificial intelligence, particularly artificial neural networks (ANN), has been widely adopted in data analysis and decision-support systems. In this study, the ANN-based model was developed using three input variables extracted from the image-processing workflow: total stripped area, total stripped perimeter, and number of stripped regions; the total aggregate area was not included as an input because it is inherently used to compute the stripping percentage. In parallel with the image-derived labels, expert-based stripping percentages were produced by senior personnel from two independent private laboratories specializing in asphalt and bituminous mixtures; all raters had at least 15 years of professional experience. To mitigate evaluator-dependent variability, images were reviewed on calibrated displays under consistent viewing conditions following a common guideline for identifying stripped regions, and when multiple experts assessed the same image, the final label was taken as the average of individual ratings. This design enables the ANN to predict stripping percentages for both the image-derived and expert-derived targets using the same inputs, thereby allowing a direct consistency check between assessment methods and facilitating an objective evaluation of model accuracy and reliability.

2.2. Image Processing Workflow (MATLAB)

All images were first resized to a standard resolution prior to analysis. The RGB frames were converted to HSV, and a blue-background mask was obtained by thresholding the H–S–V channels ($H \approx 0.55–0.75$, $S>0.3$, $V>0.3$). The aggregate surface mask was defined as the complement of this background mask. In parallel, the image was converted to grayscale to identify very dark and very bright pixels (black mask: intensity <90; white mask: intensity >210). Stripped regions were segmented as pixels belonging to the aggregate mask but not to black/white classes (soyulmaMask = agregaMask & ~blackMask & ~whiteMask). Small artifacts were removed using area opening (bwareaopen, 25 px). Connected components were then labeled (bwlabel), and region properties were extracted with regionprops (Area, Perimeter, Centroid). The total aggregate area was computed from the aggregate mask, while total stripped area and total stripped perimeter were obtained by summing per-region measurements. The stripping percentage was finally calculated as (total stripped area / aggregate area) × 100.

2.3. Data Exploration and Preprocessing

Prior to training the artificial neural network (ANN) model, basic statistical analyses were conducted on the datasets to examine the data structure in detail. Two distinct datasets were evaluated in this study: one based on image processing results and the other on expert evaluations. Since the only difference between the two datasets lies in the output variable, the focus of this study was placed on the dataset derived from image processing. The basic statistical analysis results of the expert-based dataset are not presented separately, as they were found to be largely consistent with the image processing dataset. The dataset consists of four main variables. The first three variables were designated as inputs for the model, representing the total stripped area (Input 1), the total stripped perimeter (Input 2), and the number of stripped regions (Input 3), respectively. The output variable represents the stripping percentage observed on the aggregate surface. Descriptive statistics for all 270 data points are presented in Table 1.

Table 1. Descriptive statistics for the dataset.

	count	mean	std	min	25%	50%	75%	max
Input 1	270	36351.09	35548.19	362.00	12414.25	26596.00	49234.50	248780.00
Input 2	270	9323.22	7122.37	257.58	3945.03	7597.17	13596.21	39218.10
Input 3	270	31.67	27.77	2.00	11.25	25.00	42.00	169.00
Output	270	38.69	25.96	0.74	19.45	36.97	52.52	99.76

Upon examining the dataset, the mean value of the total stripped area was calculated as 36,351.09 units². The standard deviation was found to be 35,548.19 units², indicating a wide distribution within the dataset. The minimum and maximum values were determined as 362.00 units² and 248,780.00 units², respectively, highlighting substantial variability in the stripped area values. The fact that the median is lower than the mean suggests that the data do not follow a perfectly normal distribution.

For the total stripped perimeter variable, the mean was 9,323.22 units, with a standard deviation of 7,122.37 units. The minimum and maximum values were 257.58 units and 39,218.10 units, respectively, indicating a wide variation across different samples. Regarding the number of stripped regions, the mean was calculated as 31.67, and the standard deviation as 27.77. The number of regions ranged from a minimum of 2 to a maximum of 169, reflecting notable differences between samples. The spread between quartile values further supports the conclusion that the data deviate from a normal distribution. For the stripping percentage, the mean was 38.69%, and the standard deviation was 25.96%. The minimum and maximum values were 0.74% and 99.76%, respectively. The variation in quartile values indicates that in some samples, a large portion of the surface was stripped, while in

others, relatively low stripping rates were observed. Overall, the dataset exhibits high variability across all variables and does not conform to a normal distribution. Descriptive statistics for the grouped data are presented in Table 2.

Table 2. Descriptive statistics for grouped data.

		count	mean	std	min	25%	50%	75%	max
Input 1	Full	30	81923.57	60604.96	4631.00	32510.00	75381.00	110555.25	248780.00
	Half	120	40895.13	30369.27	903.00	18311.25	37198.00	52759.50	128012.00
	Quarter	120	20413.92	15447.67	362.00	7719.75	18711.00	27392.25	66438.00
Input 2	Full	30	19845.30	11069.81	3080.77	9943.72	22182.67	30161.88	39218.10
	Half	120	10527.49	5631.79	660.91	4856.84	10758.01	15477.25	21398.63
	Quarter	120	5488.43	2981.38	257.58	2750.31	6053.38	8030.61	11285.87
Input 3	Full	30	66.30	45.92	4.00	33.75	64.50	89.25	169.00
	Half	120	35.59	24.14	2.00	16.00	33.50	51.25	93.00
	Quarter	120	19.09	12.98	2.00	8.00	16.50	28.25	52.00
Output	Full	30	38.67	25.79	2.41	19.41	35.82	51.06	98.83
	Half	120	38.68	25.86	1.00	21.33	36.57	52.68	99.60
	Quarter	120	38.70	26.32	0.74	18.17	37.20	51.70	99.76

On average, the total stripped area was calculated as 81,923.57 units² for full images, 40,895.13 units² for half images, and 20,413.92 units² for quarter images. Similarly, the total stripped perimeter was found to be 19,845.30 units for full images, 10,527.49 units for half images, and 5,488.43 units for quarter images. These results indicate a natural decrease in the stripped area and perimeter values as image size decreases. A notably high standard deviation of 60,604.96 units² was calculated for the full images, reflecting the wide distribution of the dataset. In terms of the number of stripped regions, the average was 66.30 for full images, 35.59 for half images, and 19.09 for quarter images. The corresponding standard deviation values were 45.92, 24.14, and 12.98, respectively, indicating greater variability in the number of regions among the full images. Significant differences between the minimum and maximum values (Full: 4–169, Half: 2–93, Quarter: 2–52) further demonstrate distinct variations among the groups. With respect to the stripping percentage, no substantial differences were observed between groups. The average stripping percentage was calculated as 38.67% for full images, 38.68% for half images, and 38.70% for quarter images. Standard deviation values were approximately 26% for all three groups, suggesting a wide range of variability. The minimum and maximum stripping percentages were found to be 0.74% and 99.76%, respectively, indicating that while some images exhibited extensive surface stripping, others showed minimal or no stripping at all. Overall, the dataset revealed wide variations across key variables and deviations from a normal distribution. However, unlike traditional statistical methods, Artificial Neural Networks (ANN) do not rely on the assumption of normality. The presence of broad or skewed data

distributions does not adversely affect the ANN learning process. Given its capacity to learn nonlinear relationships, ANN can effectively work with datasets exhibiting different types of distributions.

Furthermore, to normalize differences among variables with varying scales and improve the model's learning efficiency, data standardization was applied. The fact that both input and output variables had standard deviations close to 1 and means close to zero confirmed that the scaling process was performed correctly and effectively (Table 3). This procedure eliminates scale disparities between variables, enabling the model to generate more accurate predictions. It also allows for more equitable and consistent evaluations across different groups, thus enhancing the interpretability of the statistical results.

Table 3. Descriptive statistics after standardization.

		Inputs		Output
Mean:	8.22E-18	-2.23E-16	-2.63E-17	1.84E-16
Median:	0.27442	-0.24234	-0.24024	-0.06631
Standard Deviation:	1	1	1	1

According to the results of the normality test, the input variables were found not to follow a normal distribution. However, since Artificial Neural Networks (ANNs) are capable of learning nonlinear relationships, the lack of normality in the input data does not pose a concern in terms of model performance. On the other hand, the output variable was determined to exhibit a normal distribution ($p = 0.11565$), as shown in Figure 3.

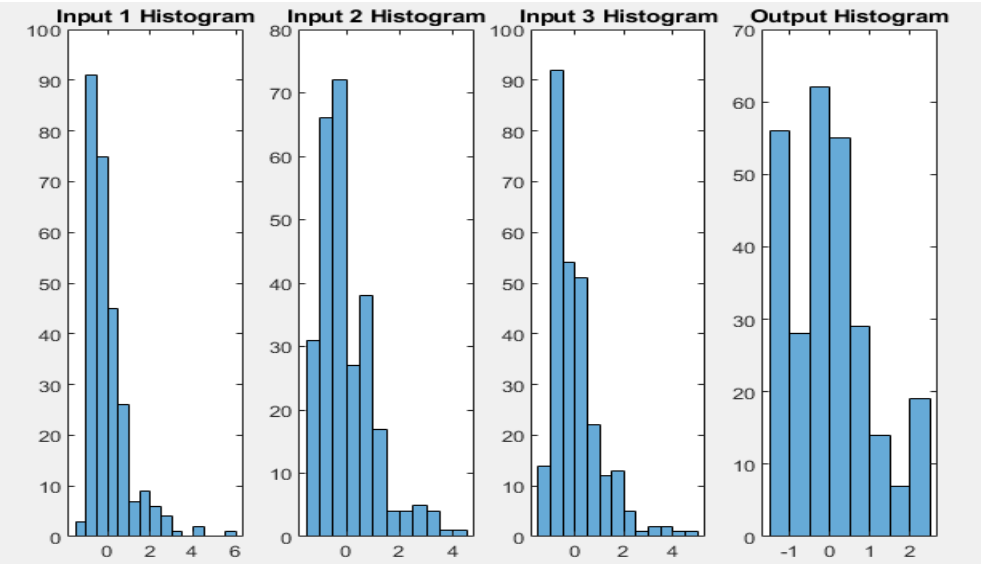

Figure 3. Shapiro-Wilk normality analysis of model variables.

The results of the correlation analysis provided a detailed examination of the relationships between the input and output variables. According to the Pearson and Spearman correlation matrices,

a strong positive correlation was observed between Input 1 and the output variable, while a moderate negative correlation was found between Input 3 and the output. Input 2 exhibited a weak positive correlation with the output variable (Figure 4).

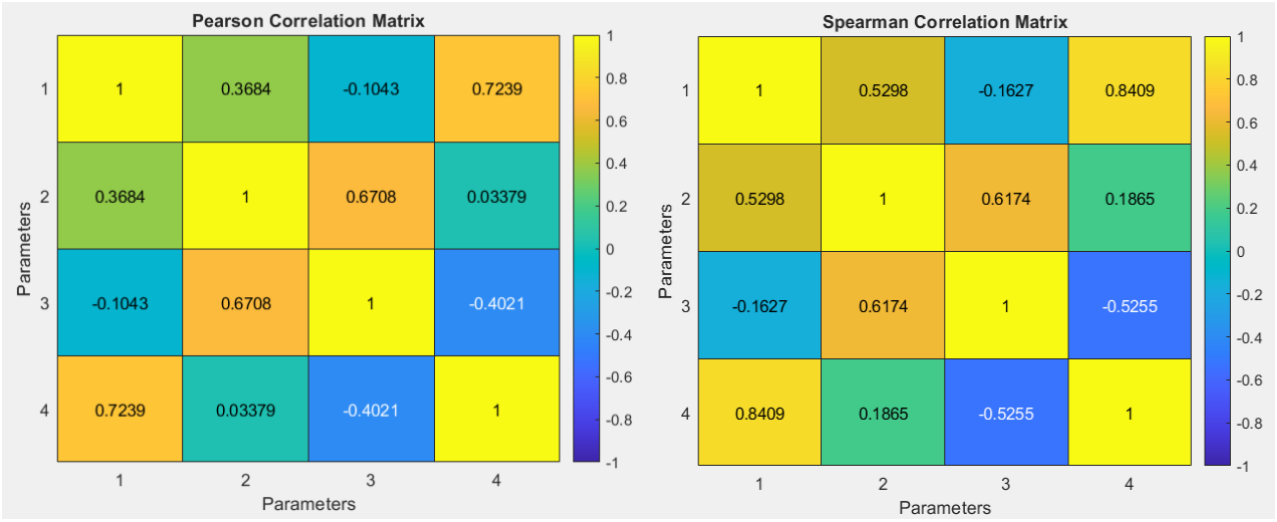

Figure 4. Pearson and Spearman Correlation Matrices Between Input and Output Variables

While correlation analysis provides insights into the relationships between variables in the context of regression models, correlation coefficients may not directly determine model performance in nonlinear approaches such as Artificial Neural Networks (ANN). Therefore, after obtaining the model results, examining the feature importance scores can offer a more accurate understanding of the impact of each variable on the output.

2.4. Model Development Using Artificial Neural Networks

In this study, artificial neural networks (ANN) were employed to develop a model for stripping prediction, utilizing the "Neural Net Fitting" tool available in MATLAB. A hidden layer with 10 neurons was configured within the application. To evaluate model performance under different configurations, various neuron counts in the hidden layer were tested. Since the number of neurons in the hidden layer directly affects the Mean Squared Error (MSE) values, this selection was made with the aim of optimizing success rates in the training, validation, and testing phases of the model. The resulting network architecture obtained upon execution of the program is presented in Figure 5.

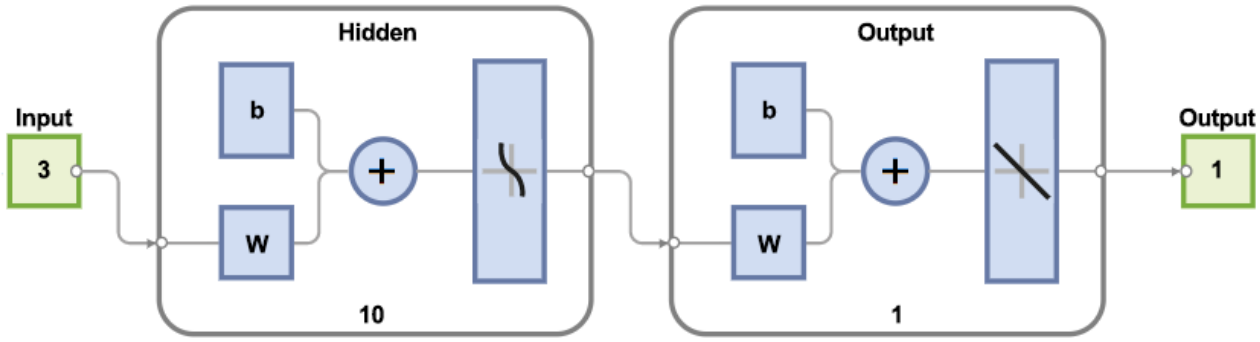

Figure 5. Architecture of the artificial neural network used in the model.

In this architecture, three input variables derived from each image interact with the 10 neurons in the hidden layer, with each connection assigned a specific weight (coefficient). Various network configurations were tested to evaluate model performance, and the architecture with 10 neurons was selected as it achieved the lowest Mean Squared Error (MSE). During the training process, the weight coefficients associated with data flow are iteratively updated and fixed upon the completion of training. These same weights are then used in the validation and testing phases to generate predictions. A similar connection structure is applied between the hidden layer and the output layer. As a result, the output value (stripping percentage) calculated during the test phase is predicted using the weight values learned during training.

The input variables of the model are the stripping area, stripping perimeter, and number of stripped regions, all determined using MATLAB. The output is defined as the stripping percentage, also calculated through MATLAB. The dataset, comprising 270 images in total, includes three input variables and one output variable for each image. Of the entire dataset, 70% (189 images) was used for training, 15% (41 images) for validation, and 15% (40 images) for testing. Model training was conducted using the Levenberg-Marquardt (LM) algorithm, an optimization method commonly employed in neural networks and often considered a variant of the backpropagation algorithm. This algorithm is particularly effective for small- to medium-sized datasets, offering fast convergence and efficient optimization of the network's weights and biases. Error plots, performance metrics, and regression values were examined in detail, and training results are presented in Table 4.

Table 4. Performance indicators of the neural network training process.

Unit	Initial Value	Stopped Value	Target Value
Epoch	0	36	1000
Elapsed Time	-	00:00:01	-
Performance	17.1	0.0409	0
Gradient	43	0.00716	1.00E-07
Mu	0.001	0.0001	1.00E+10
Validation Checks	0	6	6

The results presented in Table 4 clearly demonstrate the learning success of the artificial neural network model. During training, the performance value, used as the error metric, dropped significantly from an initial value of 17.1 to 0.0409 by the 36th epoch. This reduction indicates that the model effectively mapped the input data to the target values. The observed decrease in the gradient and the increase in the optimization parameter Mu suggest that the weight updates have stabilized and that the model has approached a local minimum. Furthermore, the validation checks performed throughout the training process indicate that the model preserved its generalization capacity and avoided overfitting. These findings confirm that the proposed neural network architecture and the selected training parameters enabled a fast and efficient learning process.

3. Findings and Discussion

This section presents the outcomes of the developed artificial intelligence models based on both image processing results and expert evaluations. It begins with a regression-based assessment of the predictions generated using image-derived data, followed by a comparative analysis of expert-based results using the same ANN architecture. Finally, the overall performance metrics of the models are evaluated and discussed to highlight the strengths and limitations of each approach.

3.1. Regression-Based Assessment of Image Processing and AI-Driven Peeling Predictions

At the end of the training phase, it was observed that the model successfully met the validation criteria. Furthermore, regression analysis was conducted to evaluate the model's prediction performance and to examine the relationship between numerical variables. Since the model includes three input variables, a multiple regression approach was adopted (Figure 6).

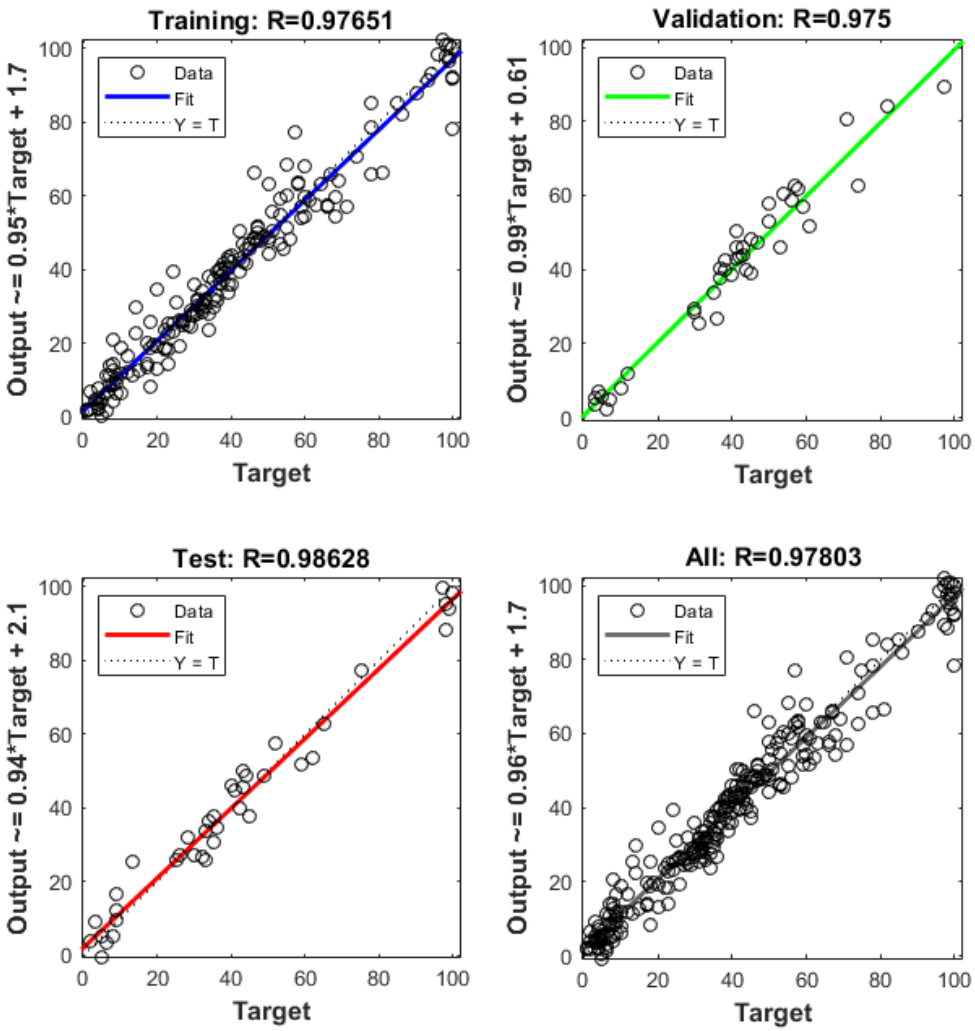

Figure 6. Regression results for stripping percentage predictions.

The linear regression plots in Figure 6 summarize the model's performance across the training, validation, and testing subsets. The regression coefficients (R) were 0.977 for training, 0.975 for validation, and 0.986 for testing. In these plots, the vertical axis denotes ANN predictions, while the horizontal axis represents the ground truth values. The tight clustering of points around the fitted lines is evidence of an effective learning process and good generalization. This, in turn, indicates that the dataset is statistically coherent and suitable for the chosen modeling strategy.

Figure 7. Training set (n = 189): actual vs. ANN predictions (image-processing).

To further illustrate sample-level behavior, Figure 7 shows the actual vs. predicted series for the 189 training images. The close agreement between traces confirms that the model reproduced the targets accurately during training. Using the parameters learned in training, we next evaluated performance on the validation set. As shown in Figure 8, predictions align closely with ground truth, demonstrating that the model sustains high accuracy beyond the training phase. Finally, testing with 40 images yielded comparable agreement (R = 0.986), reinforcing the model's robustness; these results are reported in text for brevity.

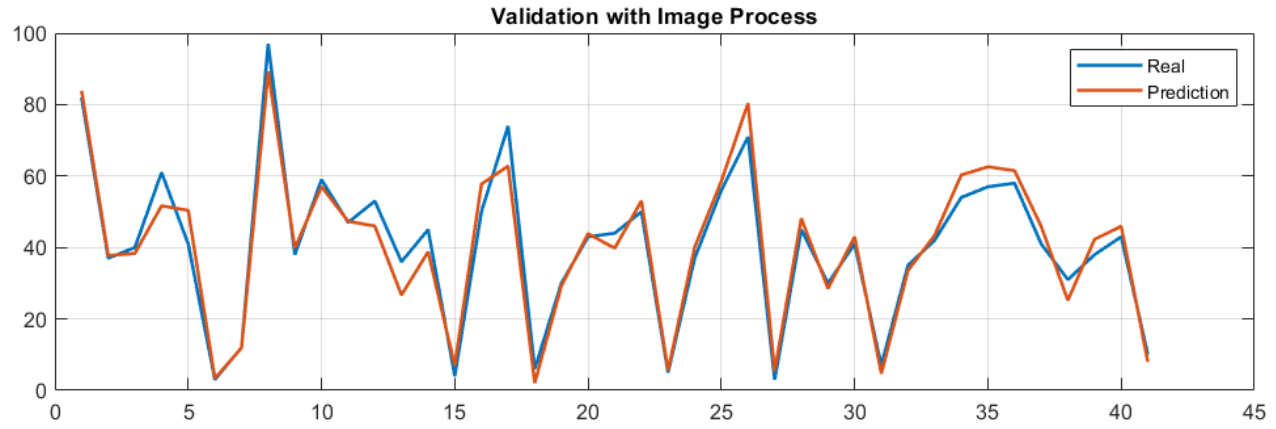

Figure 8. Validation set (n = 41): actual vs. ANN predictions (image-processing).

Upon examining Figure 8, the predicted values closely match the actual values, indicating that the model successfully completed training and maintained a high level of accuracy during validation as well.

3.2. Expert-Based Stripping Findings and Artificial Intelligence Evaluations

A similar model was developed using expert-assessed output values applied to the same input variables. In this configuration, only the outputs were changed, while the dataset, the number of inputs (three), and the processing steps remained identical to the previous model. To enable a fair comparison, the same training, validation, and test splits were used as before. After training the artificial neural network, the regression coefficients (R) were computed and are reported in Figure 9. At this stage, the R values for the training, validation, and test subsets were 0.807, 0.659, and 0.716, respectively.

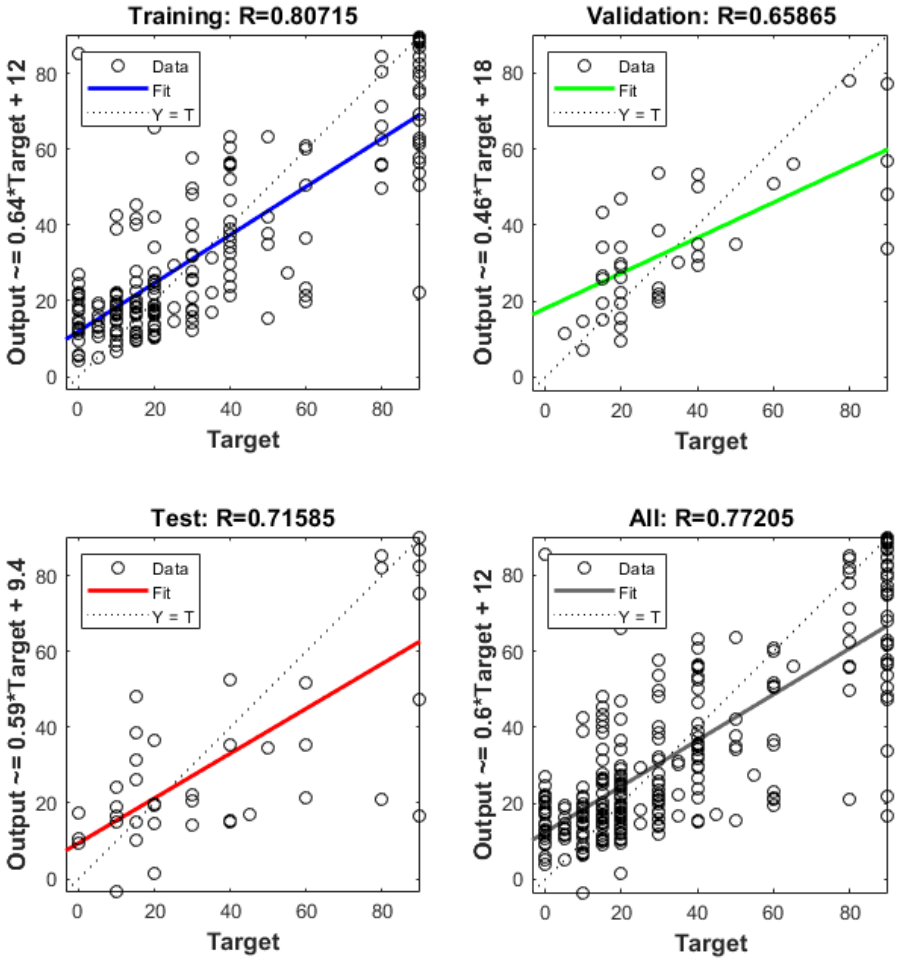

Figure 9. Regression results for expert-evaluated stripping percentages.

The discrepancy between predicted and actual values was noticeably larger than in the image-processing model. Although the fitted lines approximate the expected slope, the training process did not reach an optimal level. Consequently, more pronounced differences appeared between predictions and ground truth during testing. The limited precision of expert-based outputs and observer-to-

observer variability are considered the primary causes of this inconsistency. The actual and predicted values for the 189 training images are presented in Figure 10, while the model's performance on the 41-image validation set is shown in Figure 11. Test results (n = 40; R = 0.716) are reported in text for brevity.

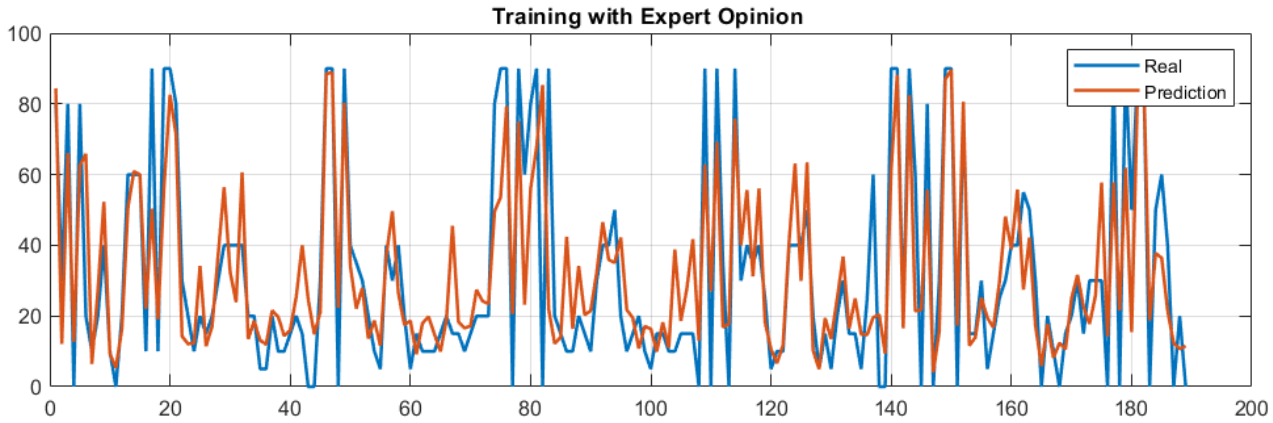

Figure 10. Actual vs. ANN-predicted stripping percentages for the training set (expert-evaluated data).

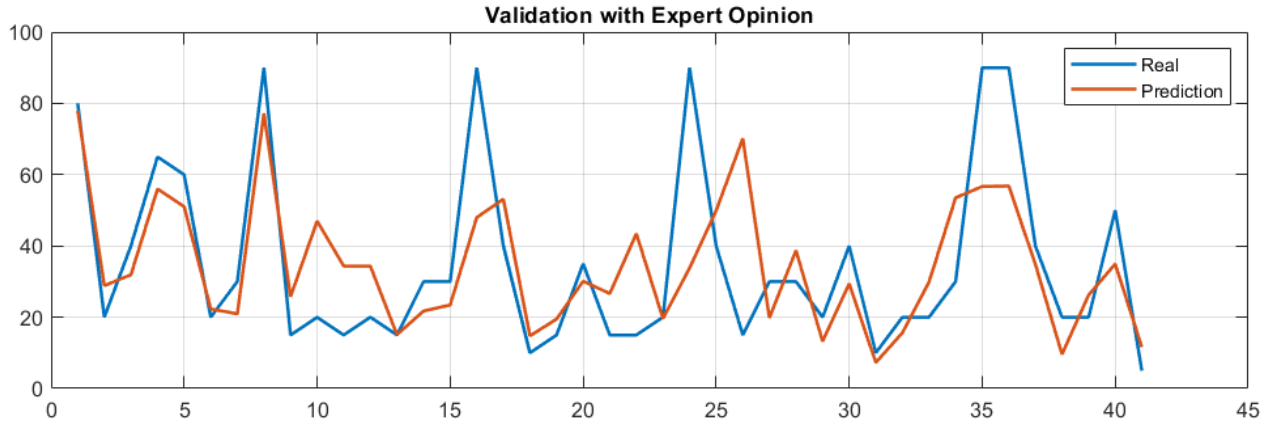

Figure 11. Actual vs. ANN-predicted stripping percentages for the validation set (expert-evaluated data).

Compared with the initial study, prediction performance in the testing phase was lower. This can be attributed to the inherently subjective nature of expert-based outputs and the difficulty of producing fully objective labels via visual assessment.

3.3. Model Performance Results

The results derived from the image processing dataset demonstrate that the artificial neural network (ANN) model exhibited high performance. The Mean Squared Error (MSE) value was as low as 0.0405, indicating a minimal error rate. Furthermore, the coefficient of determination (R^2) was calculated as 0.95935, suggesting a strong correlation between predicted and actual values. The Mean

Absolute Error (MAE) was 0.14749, further confirming that the predictions closely matched the actual values.

An analysis of the feature importance scores revealed that Input 3 (0.9639) was the most influential factor for the model, followed by Input 2 (0.8577), which also contributed significantly. In contrast, the model built on expert-based output data performed considerably worse. The MSE increased to 0.42948, and the R^2 value dropped to 0.56893, indicating a weaker predictive relationship. The MAE for the expert-based model was 0.4706, reflecting greater deviation between predicted and actual values. Feature importance analysis for the expert-based model showed that Input 2 had the highest relative importance at 0.3788. However, all input variables exhibited considerably lower importance scores, confirming the overall lower predictive power of this model. These findings clearly demonstrate that image processing data provided stronger support for the ANN model and resulted in more accurate predictions.

4. Conclusions and Recommendations

In this study, the accuracy of laboratory-based stripping tests, conducted to analyze the stripping behavior of aggregates used in road pavement structures, was evaluated, and the validity of image processing-based technological methods was investigated as an alternative. A total of 270 aggregate images obtained under laboratory conditions were analyzed using both image processing techniques and expert evaluations. A custom image processing tool developed in MATLAB was employed to calculate the stripping percentage by determining various parameters for each image, such as aggregate surface area, stripped area, stripped perimeter, and the number of stripped regions. The stripping percentages for the same samples were also assessed through visual inspection by laboratory personnel and recorded as expert opinion-based results.

According to the results, 155 out of 270 samples (57.4%) were found to be suitable based on the image processing method, whereas 203 samples (75.2%) were considered suitable according to expert evaluations. Considering that the image processing method has a higher potential to accurately estimate the stripping percentage, it was determined that expert-based assessments may include an error margin of approximately 25%. This discrepancy can be attributed to the subjective nature of expert evaluations and the inherent limitations of visual inspection methods. Overall, image processing techniques emerge as a more reliable alternative for evaluating stripping behavior, offering both objective assessment and reproducible results.

The accuracy and consistency of the obtained data were evaluated using Artificial Neural Networks (ANN) in the MATLAB environment. In the study, stripping percentages determined via both image processing techniques and expert evaluations were defined separately as output variables

in the ANN model, and two distinct analyses were conducted. Examination of the training, validation, and testing phases of the ANN revealed that the model utilizing image processing-based data exhibited higher performance, with predicted values more closely aligning with actual values. In contrast, the analysis based on expert-derived stripping percentages showed lower agreement during the testing phase. This discrepancy is attributed to the inherently subjective nature of expert evaluations and the limitations of visual inspection methods, which may lack precision in reflecting actual stripping conditions.

The findings demonstrate that image processing techniques offer a more reliable and objective alternative for evaluating stripping behavior. Notably, in both models, the feature importance score for Input 1, which represents the total stripped area, was unexpectedly low relative to expectations based on correlation analysis. This is primarily due to variations in the size of the images corresponding to different samples, which significantly affect the total stripped area. The categorization of images into three distinct groups resulted in substantial differences in total stripped area across categories, while stripping percentages remained relatively consistent regardless of image size. Therefore, the low contribution of the total stripped area to the model's predictive power is an anticipated outcome.

On the other hand, Inputs 2 and 3, namely the total stripped perimeter and the number of stripped regions, had a stronger impact on the model trained with image processing data. These results suggest that stripping behavior is not solely dependent on the total stripped area but is also influenced by the geometric characteristics of the stripped regions. Geometrical variation governs edge length and fragmentation, which in turn modulate stress concentrations and interfacial debonding pathways; hence perimeter and region count provide richer diagnostic information than area alone (Arasan, 2011). Accordingly, future protocols should complement area-based metrics with simple geometric descriptors of stripped regions (e.g., perimeter, count, circularity, size distribution).

In conclusion, the use of image processing technologies for the objective and repeatable determination of aggregate stripping rates provides more reliable and accurate results compared to traditional methods. This approach enhances measurement precision while minimizing errors associated with human subjectivity. Future research that more extensively incorporates artificial intelligence-based approaches is expected to offer significant advantages in both modeling processes and field applications.

Authors' Contributions

All authors contributed equally to the study.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The author declares that this study complies with Research and Publication Ethics.

References

Akalın, K. B., & Karacasu, M. (2020). Çevresel atıklarla modifiye edilmiş sathi kaplamaların performansının agrega-bitüm ilişkisi bağlamında değerlendirilmesi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, *32*(1), 127-136.

Aktaş, B. (2012). *Koruyucu amaçlı sathi kaplamaların performansına etki eden parametrelerin incelenmesi* Süleyman Demirel Üniversitesi]. Isparta.

Arasan, S. (2011). *Görüntü analizi ile granüler zeminlerin bazı geoteknik özelliklerinin belirlenmesi* Atatürk Üniversitesi]. Erzurum.

Beltran, G., & Romo, M. (2014). Assessing artificial neural network performance in estimating the layer properties of pavements. *Ingeniería e Investigación*, *34*(2), 11-16.

Bubeníková, E., Pirník, R., Holečko, P., & Franeková, M. (2015). The ways of streamlining digital image processing algorithms used for detection of lines in transport scenes video recording. *IFAC-PapersOnLine*, *48*(4), 174-179.

Bundy, K., Schlegel, U., Rahn, B., Geret, V., & Perren, S. (2000). An improved peel test method for measurement of adhesion to biomaterials. *Journal of Materials Science: Materials in Medicine*, *11*, 517-521.

Çetin, S. (2012). *Sathi kaplamalarda meydana gelen bozulmaların görüntü işleme yöntemiyle analizi* Süleyman Demirel Üniversitesi]. Isparta.

Doğan, G., & Ergen, B. (2022). Karayollarındaki asfalt çatlaklarının tespiti için yeni bir konvolüsyonel sinir ağı tabanlı yöntem. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, *34*(2), 485-494.

Girdap, E. (2017). *Taş mastik asfalt karışımlarda SBS ile birlikte PR plast S katkısının karışım performansı üzerindeki etkilerinin incelenmesi* Karadeniz Teknik Üniversitesi]. Trabzon.

Gürer, C., & Karaşahin, M. (2014). Sathi kaplama agregalarının adezyon özelliklerinin araştırılması. *Yapı Teknolojileri Elektronik Dergisi*, *10*(2), 1-11.

Karadağ, Ö., & Saltan, M. (2021). Bitümlü bağlayıcı ve agregaların arasındaki adezyon üzerine cocamide diethanolamide kimyasalının etkisi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, *27*(3), 312-317.

Kyo, S., Koga, T., Sakurai, K., & Okazaki, S. i. (1999). A robust vehicle detecting and tracking system for wet weather conditions using the IMAP-VISION image processing board. Proceedings 199 IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems (Cat. No. 99TH8383), Tokyo, Japan.

Lee, S., Kim, S., An, K. E., Ryu, S.-K., & Seo, D. (2018). Image processing-based pothole detecting system for driving environment. 2018 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA.

Mo, X., Sun, C., Li, D., Huang, S., & Hu, T. (2020). Research on the method of determining highway truck load limit based on image processing. *IEEE Access*, *8*, 205477-205486.

Niksaz, P. (2012). Automatic traffic estimation using image processing. *International Journal of Signal Processing, Image Processing Pattern Recognition*, *5*(4), 167-174.

Oh, I., & Barham, W. (2011). The application of artificial neural network for the prediction of the deformation performance of hot-mix asphalt. In *Computing in Civil Engineering (2011)* (pp. 227-233).

Öner, J. (2020). Seramik Atıklarıyla Hazırlanan Asfalt Karışımların Soyulmaya Karşı Dayanımının Belirlenmesi [Stripping Resistance Evaluation of Asphalt Mixtures containing Ceramic Wastes]. *Afyon*

Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, *20*(3), 498-505. https://doi.org/10.35414/akufemubid.665232

Rajab, M. I., Alawi, M. H., & Saif, M. A. (2008). Application of image processing to measure road distresses. *WSEAS Transactions on Information Science Applications*, *5*(1), 1-7.

Sajjadi, B. A. (2016). *Bazı katkıların bitümün yapışma karakteristiklerine etkisinin belirlenmesi* Süleyman Demirel Üniversitesi]. Isparta.

Sarikan, S. S., Ozbayoglu, A. M., & Zilci, O. (2017). Automated vehicle classification with image processing and computational intelligence. *Procedia Computer Science*, *114*, 515-522.

Shen, G. (2016). Road crack detection based on video image processing. 2016 3rd International Conference on Systems and Informatics (ICSAI), Shanghai, China.

Simons, S., Rossetti, D., Pagliai, P., Ward, R., & Fitzpatrick, S. (2005). The relationship between surface properties and binder performance in granulation. *Chemical engineering science*, *60*(14), 4055-4060.

Sollazzo, G., Fwa, T., & Bosurgi, G. (2017). An ANN model to correlate roughness and structural performance in asphalt pavements. *Construction Building Materials*, *134*, 684-693.

Valentova, T., Altman, J., Valentin, J., & Hamzah, M. O. (2015). Impact of Ageing and the Stability of Adhesion Additive on Moisture Susceptibility and Adhesion. *Applied Mechanics Materials*, *802*, 309-314.

White, T. D., Haddock, J. E., & Rismantojo, E. (2006). *Aggregate Tests for Hot-Mix Asphalt Mixtures Used in Pavements*. NCHRP Report.

Wigan, M. (1992). Image-processing techniques applied to road problems. *Journal of transportation engineering*, *118*(1), 62-83.

Yang, Z., Ni, C., Li, L., Luo, W., & Qin, Y. (2022). Three-stage pavement crack localization and segmentation algorithm based on digital image processing and deep learning techniques. *Sensors*, *22*(21), 8459.