Enhancing Road Surface Classification with Hybrid Deep Learning Models: A ResNet-MobileNet Approach

Estella Roberts (23102324)
Department of Computer Science
University of Galway
e.roberts10@universityofgalway.ie

Jason Winston Martin (23101224)
Department of Computer Science
University of Galway
j.martin29@universityofgalway.ie

Merlin Moncy (23103030)
Department of Computer Science
University of Galway
m.moncy1@universityofgalway.ie

Sreelakshmi Anjilimarayil Shyjumon (23103188)
Department of Computer Science
University of Galway
s.anjilimarayilshyjumon1@universityofgalway.ie

Sharaj Jagadeesan (23100668)
Department of Computer Science
University of Galway
s.jagadeesan1@universityofgalway.ie

***Abstract*—In the field of intelligent transportation systems, identifying road surfaces is crucial for improving traffic safety and streamlining vehicle control algorithms. This study presents a novel approach to road surface classification using recent developments in deep learning techniques. Motivated by the latest developments in convolutional neural network (CNN) architectures, a hybrid model is proposed that combines features from ResNet and MobileNet to improve the robustness and classification accuracy of road surfaces. By incorporating advanced techniques, we improve the model's capacity to distinguish between different types of road surfaces in various environmental circumstances. An extensive dataset covering a broad spectrum of road surface textures and appearances is used to train and assess the model. The results of our experiments indicate significant gains in classification performance over baseline models, demonstrating the effectiveness of our method in tackling road surface recognition challenges. Additionally, a user-friendly web application has been created and deployed to facilitate the practical application of the trained model for classifying road surfaces. This study enhances road surface recognition technologies and provides road organizations and stakeholders in the automotive sector with valuable insights and solutions.**

***Index Terms*—Road Surface Classification, Deep Learning Techniques, ResNet, MobileNet.**

I. Introduction

In modern transportation systems, road surface recognition is essential for controlling, navigating, and maintaining vehicle safety. Accurate surface type identification is crucial to maximizing driving assistance technologies, improving road maintenance tactics, and raising overall road safety. Traditional road surface recognition technologies are limited in their ability to handle complicated real-world circumstances since they frequently rely on manual inspection or oversimplified rule-based algorithms.

In recent years, computer vision has revolutionized due to the introduction of deep learning techniques, especially convolutional neural networks (CNNs). This has allowed notable breakthroughs in object recognition, image categorization, and semantic segmentation. By utilising CNNs, researchers have investigated a range of methods for recognising road surfaces to create reliable and practical models that can correctly identify distinct road surface types in a variety of environmental settings [1].

This report proposes a unique method for recognizing road surfaces that capitalises on the advantages of contemporary CNN designs, consisting of ResNet and MobileNet. Our methodology is driven by the necessity for a flexible and comprehensive solution that can handle the difficulties presented by real-world road surface detection jobs [2].

This paper follows a structured outline where we will first provide a brief overview of relevant literature, emphasizing essential publications that have shaped our approach. We then detail our suggested model architecture, explaining how ResNet and MobileNet features are integrated and the reasoning behind its design. After that, we provide a detailed description of the dataset we used in our tests, relevant statistics, possible preprocessing processes, and any abnormalities we found in the data. The experimental findings and a comprehensive analysis of the model's performance are presented in the following sections. Lastly, we briefly summarize our results and discuss possible directions for further road surface recognition research [3].

II. Literature Review

This section briefly overviews important studies that have substantially contributed to the development of road surface classification technology.

The development of convolutional neural network (CNN) architectures has significantly impacted the field of road surface recognition. Among these contributions, J. Balcerek et al.'s article, "Classification of road surfaces using convolutional neural network," is prominent [4]. This work tackles the crucial challenge of classifying road surfaces, which is necessary for autonomous vehicles and driver-aid systems. Balcerek et al. use visual data to classify different types of road surfaces using CNNs. Their model can efficiently distinguish between road markers and textures by extracting complex information from photos, improving road monitoring and navigation systems.

However, many challenges still exist in CNN-based road surface recognition. The variation in road surface conditions brought on by weather, road maintenance, and lighting presents an enormous obstacle. These differences may decrease model performance and reliability by adding ambiguity and complexity to the classification task. A further difficulty is the lack of relevant and diverse training datasets; getting labeled data for every situation involving a road surface can be labor and resource-intensive. Therefore, novel strategies and resilient model architectures that can efficiently manage a range of road surface conditions are necessary to tackle these obstacles.

In addition to the work by Balcerek et al., a study by Suraji et al. [5], "Identification of Road Surface Defects Using Multiclass Support Vector Machine," is another noteworthy research into the classification of road surfaces. The primary goal of this study is to use machine learning methods, particularly multiclass support vector machines (SVM), to detect surface imperfections in roads. Using Gabor filters and energy values to extract features, the authors showcase the efficacy of their approach in precisely categorizing a range of road surface flaws, including cracking, corrugation, depression, and potholes.

Although the study offers insightful information about using SVM to identify road surface defects, issues with data collecting, feature extraction, and model robustness still need to be addressed. Tackling these obstacles is crucial to improving the dependability and practicality of SVM-based methods for road surface classification.

Recent advancements in Deep Learning can be leveraged to perform tasks such as image classification more effectively than traditional methods such as SVM.

In the paper by Marcus Nolte et.al., 2018 [6], the classification of road surfaces was carried out using ResNet50 and InceptionV3 models. Images were added from Google image search to make up for the imbalance. Early stopping was applied here to prevent overfitting, and data augmentation has been carried out for better training. The InceptionV3 model terminated after 7-10 epochs, while the ResNet50 model took more time to train, with the training terminating after 10-20 epochs. While the model performs consistently during training, the different behavior on the test set, which contained the Google images for all classes, revealed that the additional data augmentation did not improve the performance and may lead to overfitting. Partial addition of Google images to the dataset was carried out to balance the classes, which resulted in increased accuracy in both models.

In contrast, when pictures for each class were added to the base dataset from Google, both the ResNet and Inception architectures exhibited overfitting behavior. Both the results highlight the importance of balance in the dataset. The well-balanced dataset resulted in an increased average classification accuracy of 92%, while the imbalanced dataset with all the classes extended with Google images resulted in an accuracy of 80%.

The possibility of automatic feature extraction and classification in Convolutional Neural Networks can be efficiently used in the image classification task carried out in the study by Tong Zhao et al., 2023 [7]. The paper proposes a decision-level fusion algorithm to classify the road surface images. The dataset's size was significant, enabling robust model training and evaluation. EfficientNetB0 was chosen as the base model in this study and combined with transfer learning. Adam optimizer is used for optimization with the learning rate decaying exponentially to ensure faster training and stable convergence. The paper uses different training strategies, such as center loss, label smoothing, data augmentation, etc., to improve the model's performance. These strategies resulted in an accuracy of 92.05% on the test set, showcasing the model's ability to generalize to unseen data. The proposed algorithm is a decision-level fusion algorithm that integrates the predictions from multiple image patches. This improved accuracy of 97.50% on the fusion test set.

III. Proposed Model

An article titled "A Comprehensive Implementation of Road Surface Classification for Vehicle Driving Assistance: Dataset, Models, and Deployment" detailed a study of road surface classification. The study classified nine image classes using CNN [7]. We found this study similar to the case study we were tasked with and has inspired the creation of a fusion method to enhance predictive performance.

To increase the predictive accuracy of road surface image classification compared to pre-trained models, a hybrid neural network was created using ResNet50 and MobileNetV3 Small. The hybrid neural network architecture combines the feature extraction capabilities of these two pre-trained models. MobileNetV3 Small is lightweight yet efficient due to its ability to extract features with fewer parameters, which helps conserve computational costs. ResNet50 is a multifaceted model known for its exceptional performance on various image classification tasks and its ability to reduce the vanishing gradient problem within deep networks.

In creating the hybrid model, the top layer used for classification in both pre-trained models (MobileNetV3 and ResNet50) has been excluded to generate an output of a feature map instead of a prediction. Feature maps were created as a form of transfer and ensemble learning to adapt and leverage the strengths of each model to increase overall predictive performance. Each layer within the pre-trained models was renamed to create unique layer names and alleviate naming

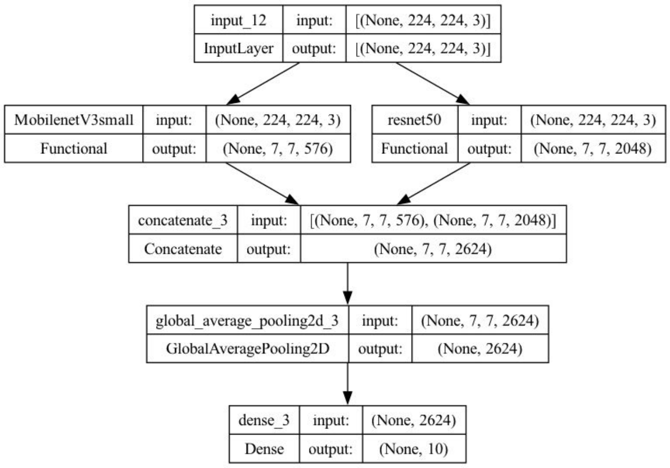

Fig. 1. Enter Caption

Fig. 2. Wet asphalt smooth

conflicts. The feature maps from both pre-trained models are then concatenated along the channel dimension to combine the extracted information into a fused compilation. A GlobalAveragePooling2D layer reduces the dimensionality of the concatenated feature maps by calculating the average value for each channel and condenses the spatial data into a flat vector. This reduction is essential to reduce the information the model needs to process to make decisions. The output is then passed into the next layer which is a Dense layer with 1024 neurons and relu as the activation. The final classification is a Dense layer containing the number of classes (3) and a Softmax activation. The model architecture of the same is shown in Fig 1. The model is then compiled using an Adam optimizer with a learning rate of 0.001 and a categorical cross-entropy loss function.

Fig. 3. Wet concrete smooth

Model	Total Parameters	Trainable Parameters	Non-trainable parameters
Resnet 50	25688963 (98.00 MB)	2101251 (8.02 MB)	23587712 (89.98 MB)
MobileNet V3	1533043 (5.85 MB)	593923 (2.27 MB)	939120 (3.58 MB)
Hybrid MobileResNet (Balanced DataSet)	**24534707 (93.59 MB)**	24469475 (93.34 MB)	65232 (254.81 KB)

TABLE I
COMPARISON OF MODEL PARAMETERS FOR INCEPTION V3, MOBILENET V3, AND HYBRID MOBILERESNET ON BALANCED DATASET

IV. DATASET AND PARTITIONING

Fig. 4. Wet gravel

The given dataset contains images from 3 different classes of road surfaces: 1. Wet asphalt smooth(Fig 2), 2. Wet concrete smooth(Fig 3), and 3. Wet gravel(Fig 4). The given dataset is a subset of the Road Surface Classification Dataset (RSCD), containing over 1 million images and over 27 different classes. The data distribution for this subset is shown in Fig. 5 and TABLE II.

Amongst the three classes, wet gravel has a coarser texture and is very different from the wet smooth concrete and wet smooth asphalt classes. Therefore, it is intuitively easier to classify than the other classes. Even between the concrete and asphalt classes, there is a subtle difference in the intensity of colors, which the model might be able to use to classify

Purpose	Total	%	Class	Number per class	%
Train	182874	97.3143	Wet Asphalt Smooth	79404	43.42
			Wet Concrete Smooth	66955	36.613
			Wet Gravel	36515	19.967
Validation	2460	1.309061	Wet Asphalt Smooth	820	33.33
			Wet Concrete Smooth	820	33.33
			Wet Gravel	820	33.33
Test	2587	1.376642	Wet Asphalt Smooth	78	3.0151
			Wet Concrete Smooth	159	6.1461
			Wet Gravel	2350	90.839

TABLE II
COMPARISON OF THE BASELINE PERFORMANCE OF DIFFERENT MODELS ON BALANCED AND IMBALANCED DATASETS

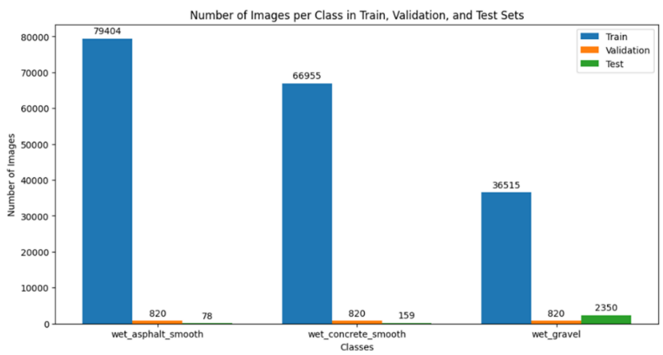

Fig. 5. Partitioning of dataset into train, validation and test sets

the images. However, noises in the image, such as water, rubble, potholes, and poor lighting, pose a challenge even to the human eye. Moreover, the dataset is imbalanced in all the classes(43:37:20). Therefore, an effective method must be devised to handle the imbalance in the dataset.

V. Experimental Results & Analysis

We carried out training on the dataset with four pre-trained models and a hybrid model. The pre-trained models consist of EfficientNet B0, ResNet 50, Inception V3, and MobileNet V3. These four pre-trained models were referenced in a study of road image classification [comprehensive]. We used these models to train and test on a dataset sample to create a baseline for reference. We have tested all pre-trained models on a balanced subset of the dataset consisting of 10,000 images per class, totaling 30,000 images. Given the lack of computational resources, we used 30,000 images at the maximum. Therefore, we decided to split the training dataset into 10,000 images for each class, although we completely understand that we almost always deal with imbalanced datasets in real-world scenarios. We also decided to train the model on 15

We chose the two highest-performing models based on accuracy and created a hybrid model. The two highest-performing pre-trained models were ResNet 50 and MobileNet, which were used to develop our hybrid MobileResNet model. Combining these two models significantly increased the total trainable parameters, resulting in a better loss reduction over epochs, thus increasing accuracy. The hybrid MobileResNet model outperformed all others, obtaining a test accuracy of 97.75%. With this increase in accuracy, we demonstrate its ability to detect road surfaces in various scenarios correctly. The hybrid method proved to be effective, retaining a high accuracy of 95.70% even when dealing with an imbalanced dataset.

Model	Test Accuracy (%)	Loss
EfficientNet	78.58	0.74
ResNet50	84.29	0.933
Inception V3	62.79	0.449
MobileNet V3	82.61	0.628
Hybrid MobileResNet (Balanced DataSet)	**92.31**	0.227
Hybrid MobileResNet (Imbalanced DataSet)	**92.15**	0.922

TABLE III
COMPARISON OF THE BASELINE PERFORMANCE OF DIFFERENT MODELS ON BALANCED AND IMBALANCED DATASETS (10 EPOCHS)

Model	Test Accuracy (%)	Loss
EfficientNet	88.56	0.369
ResNet50	92.88	0.636
Inception V3	68.95	0.635
MobileNet V3	90.17	0.788
Hybrid MobileResNet (Balanced DataSet)	**97.75**	0.102
Hybrid MobileResNet (Imbalanced DataSet)	**95.70**	0.183

TABLE IV
COMPARISON OF THE BASELINE PERFORMANCE OF DIFFERENT MODELS ON BALANCED AND IMBALANCED DATASETS (100 EPOCHS)

To ensure effective convergence during the training phase, the models were trained using an Adam optimizer with a learning rate of 0.001. Over the course of 100 epochs, we experimented with several model architectures and training techniques to find the most efficient strategy for classifying road surfaces.

The front end for the hybrid model created in this case study is a web application developed using the Python library Streamlit. The web application has been deployed on the Internet using the Streamlit Cloud and is accessible through the following URL: https://roadsurfaceclassification.streamlit.app/. This interface allows the user to add an image and the model classifies the image into one of the 3 classes. The user can change the model to test different models on the same image to compare the confidence of each model's prediction.

VI. Performance Analysis

A. Accuracy and Loss Curves

The accuracy and loss graphs for the pre-trained models EfficientNetB0, ResNet50, InceptionV3, and MobileNetV3 exhibit a considerable amount of fluctuation over the epochs. Ideally, we would expect to see consistent improvement in accuracy, but these models experience significant fluctuations with high variance that actually result in poorer performance compared to previous epochs. The hybrid model's accuracy shows steady improvement until approximately the 16th epoch before becoming more variable. However, these fluctuations are not as pronounced as those seen in the pre-trained models. Furthermore, there is less disparity between train and test accuracy for this model compared to others, which suggests reduced overfitting.

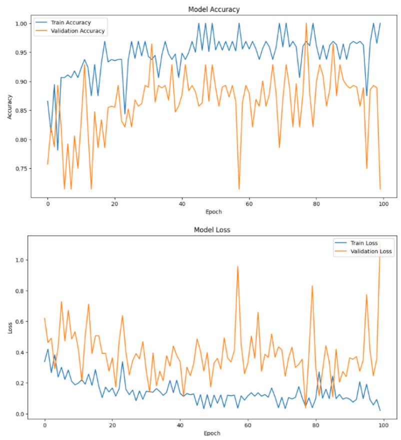

Fig. 6. EfficientNet Model

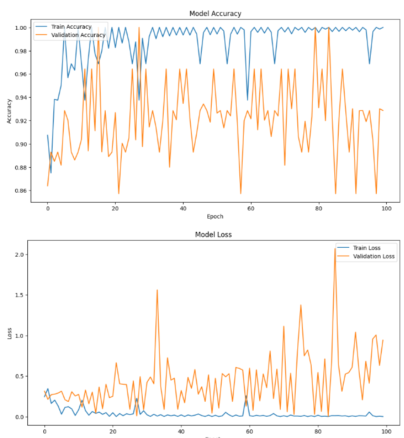

Fig. 7. ResNet 50

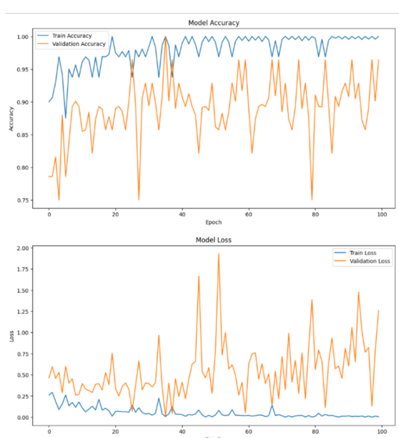

Fig. 9. MobileNet V3

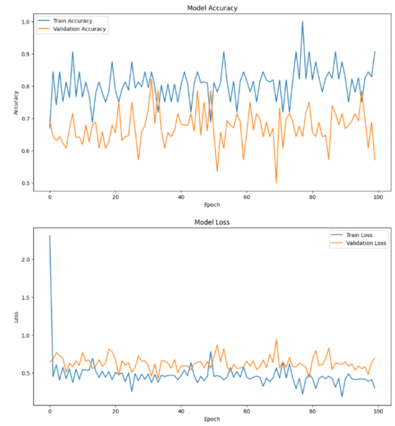

Fig. 8. Inception v3

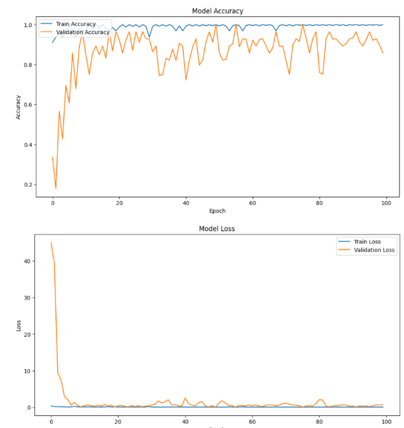

Fig. 10. Hybrid-MobileResNet

B. *Confusion Matrix*

In this confusion matrix for ResNet50, the diagonal numbers represent the true positives for each class (267, 302, 271). This model seems to classify the majority of classes well, but we are seeing this model encountering difficulty when classifying the wet concrete class. The highest misclassification within this model was classifying wet asphalt smooth as wet concrete smooth

In this confusion matrix for MobileNet, the diagonal numbers represent the true positives for each class (196, 321, 242). The miss classification of the wet asphalt smooth class and the wet gravel class increases compared to the results presented in the ResNet 50 confusion matrix. Wet gravel proved to be the most difficult for this model to classify as wet concrete smooth was the most successfully classified class. The highest misclassification in this model proved to be wet gravel as wet concrete smooth.

In this confusion matrix for the hybrid model, the diagonal numbers represent the true positives for each class (243, 314, 263). Compared to MobileNet, the misclassification of wet concrete has decreased but is still slightly higher than the misclassification of this class in Resnet. Similarly to the mobileNet confusion matrix, the highest misclassification occurring is wet gravel as wet concrete smooth. Wet concrete is the most correctly classified class in all three models.

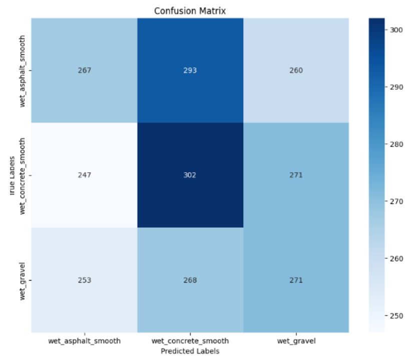

Fig. 11. ResNet 50

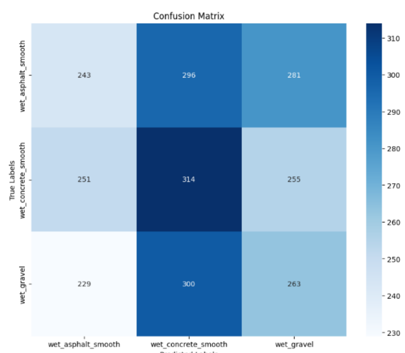

Fig. 12. Hybrid-MobileResNet (Balanced Dataset)

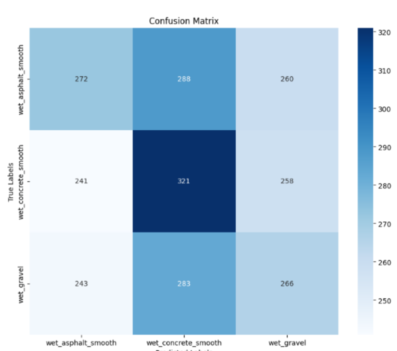

Fig. 13. Hybrid-MobileResNet (Imbalanced Dataset)

Conclusion and Future Work

The hybrid model presented in this work offers a novel approach to the age-old image classification problem. While traditional pre-trained models provide good results, the performance of two models combined in an ensemble arrangement is more effective. Higher accuracy is achieved, but at the cost of extra computational power, particularly while utilizing a smaller subset of the dataset. Despite the increased computational requirements, the model displays higher accuracy, demonstrating its potential for use in various fields, such as autonomous car systems, road infrastructure evaluation, and anomaly detection in transportation networks, such as locating potholes, etc. The user can easily modify the hybrid model to fit various models, making it possible to customize the hybrid model to the particular requirements of the problem or use case at hand and ensure scalability at all times. For future improvements, additional steps could be included to increase the classification accuracy such as preprocessing steps to enhance the features within the images, increasing neurons within the neural network, and parameter optimization.

References

[1] Oladimeji, D.; Gupta, K.; Kose, N.A.; Gundogan, K.; Ge, L.; Liang, F. Smart Transportation: An Overview of Technologies and Applications. Sensors 2023, 23, 3880. https://doi.org/10.3390/s23083880

[2] Li, X.; Xu, X.; He, X.; Wei, X.; Yang, H. Intelligent Crack Detection Method Based on GM-ResNet. Sensors 2023, 23, 8369. https://doi.org/10.3390/s23208369

[3] Ganguly , Abhisek , et al. Evaluating CNN Architectures Using Attention Mechanisms: Convolutional Block Attention Module, Squeeze, and Excitation for Image Classification on CIFAR10 Dataset. 14 Aug. 2023.

[4] J. Balcerek, A. Konieczka, K. Piniarski and P. Pawłowski, "Classification of road surfaces using convolutional neural network," 2020 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), Poznan, Poland, 2020, pp. 98-103, doi: 10.23919/SPA50552.2020.9241254.

[5] A. Suraji, A. Tugas Sudjianto, R. Riman, C. Aditya, R. Pahlevi Putra and A. Yuniar Rahman, "Identification of Road Surface Defects Using Multiclass Support Vector Machine," 2023 4th International Conference on Artificial Intelligence and Data Sciences (AiDAS), IPOH, Malaysia, 2023, pp. 5-10, doi: 10.1109/AiDAS60501.2023.10284716.

[6] M. Nolte, N. Kister and M. Maurer, "Assessment of Deep Convolutional Neural Networks for Road Surface Classification," 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 2018, pp. 381-386, doi: 10.1109/ITSC.2018.8569396.

[7] T. Zhao, J. He, J. Lv, D. Min, and Y. Wei, "A Comprehensive Implementation of Road Surface Classification for Vehicle Driving Assistance: Dataset, Models, and Deployment," in IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 8, pp. 8361-8370, Aug. 2023, doi: 10.1109/TITS.2023.3264588.