2025 IEEE 25th International Symposium on Computational Intelligence and Informatics (CINTI) | 979-8-3315-5291-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/CINTI67731.2025.11311744

A Feature-Level Approach for Outdoor Surface Classification

1st Erdal Alimovski
Computer Engineering Department
Istanbul Sabahattin Zaim University
Istanbul, Türkiye
erdal.alimovski@izu.edu.tr

2nd Asiye Demirtaş
Electrical-Electronics Engineering Department
Istanbul Sabahattin Zaim University
Istanbul, Türkiye
asiye.demirtas@izu.edu.tr

3rd Yağmur Uşak
Marine Engineering Department
Istanbul Technical University
Istanbul, Türkiye
usakyagmur1@gmail.com

***Abstract*— In this study, we address the problem of outdoor surface classification using deep convolutional neural networks. A custom dataset was generated to represent various outdoor ground types under different conditions. State-of-the-art CNN models, including VGG19, ResNet50, and InceptionV3, were employed and evaluated individually on this dataset. Based on the comparative results, we proposed a feature-level fusion model that combines VGG19 and InceptionV3 to leverage their complementary strengths. Experimental results show that the proposed fusion model significantly outperforms the individual models, achieving 97% in precision, recall, F1-score, and test score. These findings demonstrate the effectiveness of the ensemble approach in improving classification performance for outdoor surface recognition tasks.**

Keywords— surface, classification, cnn, robotics.

I. INTRODUCTION

In recent years, autonomous mobile robotic systems have became increasingly important in fields such as millitary operations [1], search and rescue [2], transportation [3] and agriculture [4]. As robots are being used more widely, they have gradually started to replace humans to complete certain tasks. However, the robots are sensitive to environments, making its functionality difficult in certain settings. In this case, knowledge about the surface type becomes crucial for ensuring stable and effective operation.

At the present, surface classification in mobile robots typically are performed utilizing onboard sensors in order to recognize both present and upcoming surface. The interaction between sensors and surface can be categorized into two approaches: non-interactive and interactive surface classification [5]. Non-interactive surface classification methods in general are implemented by optical sensors such as digital cameras, depth cameras, and laser imagin detection and ranging (LiDARs) [6, 7]. In other hand, interactive methods performs surface classification through sound, tactile, and vibration signals generated during the interaction between the mobile robot and the ground [8, 9, 10].

Computer vision and Deep Learning (DL) based techniques have played a significant role in diverse applications. By extracting high level features from visual data, these approaches enable more accurate and flexible decision-making in complex environments. While DL and camera based sensing are commonly employed, the literature also include studies based on interactive methods. Vicente et.al [11] collect vibration signals utilizing inertial measurement unit (IMU) sensors mounted on a mobile robot equipped with four omnidirectional wheels. Feature extraction was performed in both the time and frequency domain, followed by dimensionality reduction through principal component analysis (PCA). For classification of four different surface types extreme learning machine algorithm were carried out, and an accuracy of 85% was achieved. In [12], the authors proposed a three-dimensional surface classification method. Vibration data were obtained from sensor placed on a Clearpath Jackal robot, and features were extracted using Fast Fourier transform followed by normalization. A multi-layer perceptron (MLP) network was trained to map the relationship between vibration signals and terrain types. Proposed method was evaluated accros different environments and speeds. The classification accuracy ranges between 85-95%. Authors in [13] proposed lightweight surface classification method based on time-domain features from inertial and magnetic sensors. The proposed method employs MLP for classification and is tested on diverse outdoor surfaces. Obtained results indicate that gyrospoce data outperforms accelerometer data for classification task, while magnotometer data alone is inadequate but enhanced the overall performance when combined with IMU sensor. In [14], method based on a combionation of visual and proprioceptive signals was proposed for surface classification in low-light challenging outdoor environments. The proposed method is composed of CNN which utilizes IMU and encoder signals. In addition, it is enhanced with fuzzy-tuned Plateau Equalization process to handle low-light conditions. Experimental results shows that proposed method achieves remerkable performance in challengin environments with low light.

In this paper, we employ state-of-the-art CNN models, namely VGG19, ResNet50, and InceptionV3, for outdoor surface classification on a custom generated dataset. In addition, based on the individual performances of the models, we propose a feature-level fusion approach that combines InceptionV3 and VGG19 to enhance the classification accuracy.

The rest of the paper is orginized as follows: Section II describes the dataset, employed models and proposed approach. Section III demonstrates the experimental results and section IV concludes the paper.

II. Materials and Methods

A. Dataset

Due to the lack of open source outdoor surface dataset, we generated a new dataset comprising three different surface types: asphalt, concrete, and grass. The samples were collected in various outdoor environments during different times of the day, including morning, afternoon and late afternoon hours, to capture variations in natural lighting conditions. Furthermore, data was primarily collected using four different smartphone models to ensure variation in camera characteristics. All samples in the dataset were manually annotated to ensure accurate labeling of the surface types. To support reproducibility and further research, the dataset will be made available to interested researchers upon request. The dataset include 438 images of asphalt, 464 of concrete and 500 of grass, totaling 1402 samples. Some samples of the dataset are sohown in Figure 1. All the collected samples are in RGB format. Altought the image size of samples vary, all samples were resized to a equivalent dimension in order to fed into the models. The dataset was split into training, validation, and testing sets, with 80% used for training, 10% for validation, and the remaining 10% for testing. Table I provides detailed overview of the dataset.

asphalt concrete grass

Fig. 1. Some samples of generated dataset.

TABLE I. Details of the generated dataset.

Dataset	Asphalt	Concrete	Grass	Total
Train	400	382	399	1181
Validation	38	40	40	118
Test	45	42	61	148
Total	438	464	500	1402

B. VGG19

VGG19 is a deep CNN architecture, introduced by the Visual Geometry Group of Oxford University utilized for image recognition tasks. With its remerkable performance in the 2014 ImageNet competition, VGG19 emerged as a benchmark deep learning architecture. Its architecture consist of a total 19 layers; 16 convolutional layerss and 3 dense layers [15]. The architecture takes an input of 224x224 pixels.

Identify applicable funding agency here. If none, delete this text box.

C. ResNet50

ResNet50 is a deep CNN architecture introduced by He et. Al [16] in 2015 as part of the Residual Networks (ResNet) family. It achieved first place in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2015 by addressing the degradation problem observed in very deep networks. The total depth od the model is 50 layers, including convolutional, batch normalization and dense layers which are orginized in residual blocks. The core architecture of ResNet50 draws insipration ftom the VGG architectures, particularly in its use of stacked convolutional layers.

D. InceptionV3

InceptionV3 is enhanced version of the original Inception (V1) architecture, designed for image recognition tasks [17]. It incorporates several several architectural improvements to increase efficiency and accuracy. Developed by Google, InceptionV3 integrates techniques such as factorized convolutions where a 3×3 convolution is decomposed into 3×1 followed by 1×3 convolutions to reduce computational cost while preserving representational power. This approach allows for more efficient feature extraction and significantly decreases the number of parameters, enabling faster training.

E. Proposed Fusion Model

Extracting high level features from outdoot gorund surfaces remains challengin task due to surface texture similarity, varying lighting conditions, and occlusions. Traditional CNN architectures may struggle to generalize across such diverse environments when used independently. To overcome this limitations, we propose feature-level fusion approach that combines the strengths of VGG19 and InceptionV3 architectures. In the proposed architecture, both VGG19 and InceptionV3 are used as parallel feature extractors by removing their classification layers and freezing their pre-trained weights. For a given input x, deep features are extracted separately from both VGG19 and InceptionV3 networks, resulting in two vectors $f_{vgg19}(x) \in \mathrm{R}^{n}$ and $f_{incV3}(x) \in \mathrm{R}^{m}$. This feature vectors are then concatenated to form a single unified representation $F(x) = [f_{vgg19}(x) \,||f_{incV3}(x)] \in \mathrm{R}^{n+m}$, which is subsequently passed through fully connected layers to perform the final classification.

III. Experiments

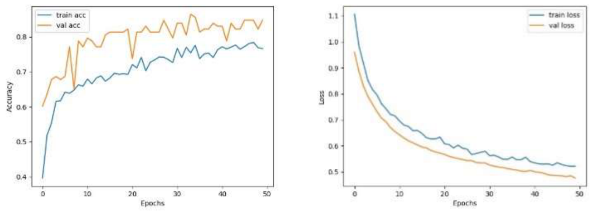
In this study, we address the surface classificaion problem by evaluation the perofmance of state-of-the-art CNN models and proposed fusion method. Training and testing were carried out on a Workstation equipped with an an AMD Ryzen 9 5900X processor, 64 GB of RAM, and an NVIDIA GeForce RTX 3080 Ti GPU. The dataset was split into 80% training, 10% validation, and 10% testing, and the models were evaluated using following metrics:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{1}$$

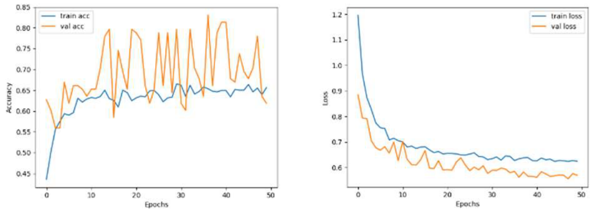
$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

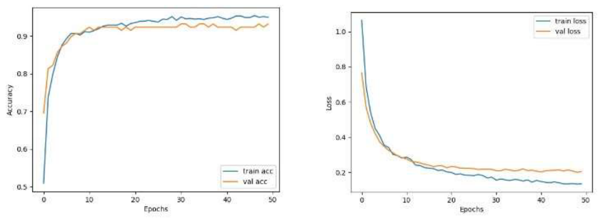
$$F1score = \frac{2\,x\,Precision\,x\,Recall}{Precision+Recall} \tag{4}$$



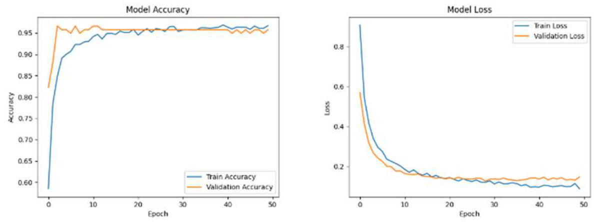
VGG19



ResNet50

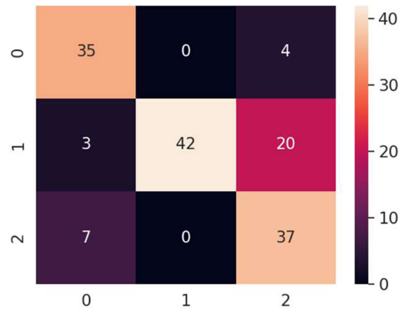


InceptionV3

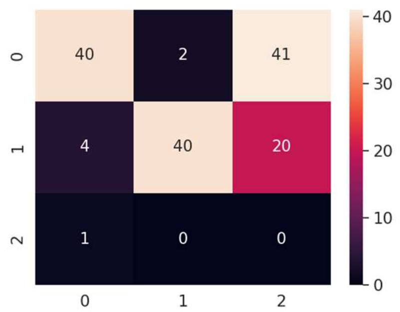


Proposed Fusion Model

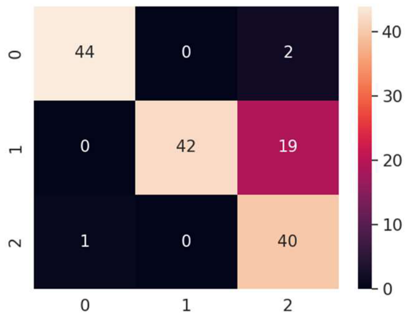
Fig. 2. Training and validation accuracy-loss graphs of each models.



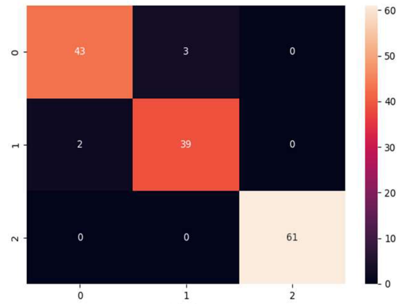
VGG19



ResNet50



InceptionV3



Proposed Fusion Model

Fig. 3. Confusion Matrices of each method.

Figure 2 presents the training and validation accuracy and loss curves for the VGG19, ResNet50, InceptionV3, and the proposed fusion model over 50 epochs. The VGG19 model shows a steady improvement in both training and validation accuracy, though a noticeable gap between them suggests moderate overfitting. In contrast, ResNet50 demonstrates unstable validation accuracy throughout the training process, with frequent fluctuations indicating inconsistent generalization performance. InceptionV3 achieves high and stable accuracy on both training and validation sets, showing strong convergence and robustness with minimal overfitting. The proposed fusion model outperforms all individual models by achieving the highest and most stable accuracy levels, with closely aligned training and validation curves. Additionally, the fusion model exhibits a rapid decrease in training and validation loss, indicating efficient learning and better generalization

Figure 3 compares the confusion matrices of VGG19, ResNet50, InceptionV3, and the proposed Fusion model. Here, Class 0 corresponds to asphalt, Class 1 to concrete, and Class 2 to grass. As illustrated in Figure 3, ResNet50 exhibits substantial misclassifications, particularly failing to correctly predict samples from the grass class (Class 2), which accounts for its low recall and F1-score. VGG19 performs moderately better but still confuses concrete and grass (Class 1 and Class 2), leading to reduced overall accuracy. InceptionV3 achieves higher precision and recall, correctly identifying most asphalt and concrete samples but still showing some misclassification between concrete and grass. In contrast, the proposed fusion model correctly classifies almost all samples, misclassifying only two asphalt samples as concrete and three concrete samples as asphalt, while perfectly identifying all grass samples. This demonstrates the model's strong robustness and generalization capability across different surface types.

TABLE II. PERFORMANCE RESULTS OF APPLIED STATE-OF-THE-ART MODELS AND PROPOSED MODEL.

Method	**Precision (%)**	**Recall (%)**	**F1Score (%)**	**Test (%)**
VGG19	79	79	77	77
ResNet50	61	37	46	54
InceptionV3	88	87	86	85
Proposed	96	96	97	97

Table II presents a comprehensive comparison of the evaluation metrics for all applied state-of-the-art models and the proposed fusion model. ResNet50 yields the weakest performance, with a test accuracy of 54% and an F1-score of only 46%, indicating poor generalization. VGG19 performs moderately with balanced precision and recall (both 79%), but still falls short in overall performance. InceptionV3 provides a marked improvement over the previous baselines, achieving 85% test accuracy, 88% precision, 87% recall, and an F1-score of 86%. These results indicate that InceptionV3 is more capable of capturing discriminative features from the dataset, improving both precision and recall. The proposed fusion model outperforms all baselines, achieving a test accuracy of 97%, precision of 96%, recall of 96%, and an F1-score of 97%. nlike the other models, it maintains both high precision and recall simultaneously, which indicates its robustness in minimizing both false positives and false negatives. The consistent and superior results across all evaluation metrics highlight the effectiveness of the fusion strategy in capturing complementary features and improving generalization. These findings confirm that the proposed model provides a reliable and balanced solution for outdoor surface classification.

TABLE III. COMPARISON OF THE PROPOSED MODEL AND EXISTING STUDIES.

Research	**Surface Type**	**Method**	**Accuracy (%)**
[11]	indoor/ outdoor	PCA+Extreme ML	85
[12]	Outdoor	FFT+MLP	85-96
[18]	indoor/ outdoor	Customized CNN model	88
[19]	indoor/ outdoor	ResNet50	80
Proposed	Outdoor	Fusion VGG19+IncV3	97

Table III provides a comparative overview of the proposed method and related studies in terms of surface type, methodology, and classification accuracy. Previous research has investigated various approaches for surface type recognition. For instance, PCA combined with traditional machine learning algorithms [11] and FFT-based feature extraction with a multilayer perceptron [12] have reported accuracies between 85% and 96% under both indoor and outdoor conditions. More recent deep learning-based approaches, such as customized CNN architectures [18] and ResNet50-based models [19], have achieved accuracies of 88% and 80%, respectively, demonstrating the potential of convolutional networks but also highlighting their limitations in generalization across diverse environments. In contrast, the proposed fusion model, which integrates VGG19 and InceptionV3 feature representations, achieves the highest accuracy of 97% in outdoor environments. This significant improvement emphasizes the model's superior capability to extract complementary and discriminative features, resulting in more robust and reliable classification of different surface types compared to previous methods.

IV. CONCLUSION

In this study, we addressed the challenge of outdoor surface classification by utilizing CNN models. A custom dataset was generated under varying environmental conditions. In order to analyze the performance of CNN models in custom generated dataset, we perform state-of-the-art CNN models such as VGG19, ResNet50, and InceptionV3. In addition, to enhance the classification accuracy, we proposed feature-level fusion model that combines the strength of VGG19 and InceptionV3. The proposed method achieved 97% across all evaluation metrics, outperforming the applied state-of-the-art models and prior studies. Obtained results confirms the effectiveness of the proposed fusion model for robust outdor surface classification. Given its high classification accuracy and generalization ability, the proposed model can be integrated into robotic systems for tasks such as autonomous navigation, surface-aware locomotion, and surface-adaptive control.

Future work will focus on applying the proposed fusion model to different open-source datasets for outdoor surface classification to evaluate its generalization ability. Additionally, the model will be implemented on a real robotic platform to test its performance in real-world scenarios

REFERENCES

[1] Nevliudov, I., Yevsieiev, V., Maksymova, S., & Chala, O. (2025). A small-sized robot prototype development using 3D printing. *acta mechanica et automatica*, *19*(1).

[2] Drew, D. S. (2021). Multi-agent systems for search and rescue applications. Current Robotics Reports, 2(2), 189-200.

[3] Hercik, R., Byrtus, R., Jaros, R., & Koziorek, J. (2022). Implementation of autonomous mobile robot in smartfactory. *Applied Sciences*, *12*(17), 8912.

[4] Baek, E. T., & Im, D. Y. (2022). ROS-based unmanned mobile robot platform for agriculture. *Applied Sciences*, *12*(9), 4335.

[5] Wang, M., Ye, L., & Sun, X. (2021). Adaptive online terrain classification method for mobile robot based on vibration signals. *International Journal of Advanced Robotic Systems*, *18*(6), 17298814211062035.

[6] Vandapel, N., Huber, D. F., Kapuria, A., & Hebert, M. (2004, April). Natural terrain classification using 3-d ladar data. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings.*

[7] Narváez, F. Y., Gregorio, E., Escolà, A., Rosell-Polo, J. R., Torres-Torriti, M., & Cheein, F. A. (2018). Terrain classification using ToF sensors for the enhancement of agricultural machinery traversability. *Journal of Terramechanics*, *76*, 1-13.

[8] Christie, J., & Kottege, N. (2016, May). Acoustics based terrain classification for legged robots. In *2016 IEEE international conference on robotics and automation (ICRA)* (pp. 3596-3603). IEEE.

[9] Valada, A., & Burgard, W. (2017). Deep spatiotemporal models for robust proprioceptive terrain classification. *The International Journal of Robotics Research*, *36*(13-14), 1521-1539.

[10] Wu, X. A., Huh, T. M., Mukherjee, R., & Cutkosky, M. (2016). Integrated ground reaction force sensing and terrain classification for small legged robots. *IEEE Robotics and Automation Letters*, *1*(2), 1125-1132.

[11] Vicente, A., Liu, J., & Yang, G. Z. (2015, September). Surface classification based on vibration on omni-wheel mobile base. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 916-921). IEEE.

[12] Bai, C., Guo, J., Guo, L., & Song, J. (2019). Deep multi-layer perception based terrain classification for planetary exploration rovers. *Sensors*, *19*(14), 3102.

[13] Sarcevic, P., Csík, D., Pesti, R., Stančin, S., Tomažič, S., Tadic, V., ... & Odry, A. (2023). Online outdoor terrain classification algorithm for wheeled mobile robots equipped with inertial and magnetic sensors. *Electronics*, *12*(15), 3238.

[14] Chen, H. Y., Sang, I. C., Norris, W. R., Soylemezoglu, A., & Nottage, D. (2024). Terrain classification method using an NIR or RGB camera with a CNN-based fusion of vision and a reduced-order proprioception model. *Computers and Electronics in Agriculture*, *227*, 109539.

[15] Zakaria, N., & Hassim, Y. M. M. (2024). A review study of the visual geometry group approaches for image classification. *Journal of Applied Science, Technology and Computing*, *1*(1), 14-28.

[16] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[17] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

[18] Singh, S., Sajwan, M., Singh, G., Dixit, A. K., & Mehta, A. (2023). Efficient surface detection for assisting collaborative robots. *Robotics and Autonomous Systems*, *161*, 104339.

[19] Kurobe, A., Nakajima, Y., Kitani, K., & Saito, H. (2021). Audio-visual self-supervised terrain type recognition for ground mobile platforms. *IEEE Access*, *9*, 29970-29979.