ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20231018184747

Original scientific paper
Received: 2023-10-18, Accepted: 2024-01-08

Improving Freedom of Visually Impaired Individuals with Innovative EfficientNet and Unified Spatial-Channel Attention: A Deep Learning-Based Road Surface Detection System

Amit Chaudhary*, Prabhat Verma

Abstract: Individuals with visual impairments often encounter substantial challenges navigating outdoor spaces due to their inability to perceive road-surface conditions. This study introduces an innovative method that harnesses deep learning to identify and categorize road surfaces, aiming to enhance the independence and mobility of the visually impaired. Leveraging the EfficientNetB0 model as a foundational framework and employing unified spatial-channel attention, we classified road surface images captured from a wearable camera. Through rigorous training and evaluation on a substantial dataset of road images, our modified system exhibited remarkable performance, accurately identifying road surfaces with an impressive 99.39% accuracy rate. This deep learning-driven approach holds promise as a pivotal tool for improving the autonomy and safety of individuals with visual challenges by providing instantaneous feedback on road conditions.

Keywords: attention mechanism; deep learning network; EfficientNet-B0; pedestrian with vision limitations

1 INTRODUCTION

The World Health Organization (WHO) states that approximately 2.2 billion individuals experience visual impairments [1]. This poses significant challenges for visually impaired individuals, who struggle to navigate and interact with their surroundings. They lack environmental information and have difficulty recognizing objects and people, limiting their independence and access to important services. To tackle this issue, various navigation solutions have been proposed that cater to different GPS-based, audio-based, and smartphone-based systems. However, GPS-based systems can be unreliable in urban areas and indoors owing to their poor signal quality [2]. Audio-based systems struggle in noisy environments [3] and do not provide object-location information [4]. Smartphone-based systems have limitations in terms of their battery life and indoor functionality. Given these obstacles, there is a growing emphasis on research to enhance the freedom and movement of those with visual impairments through deep learning-based road surface detection. By utilizing deep learning techniques to detect and classify road surfaces, this study aims to significantly enhance the lives of visually impaired individuals. This technology enables them to navigate safely and efficiently by providing real-time information about road surfaces, including detecting obstacles and changes in elevation. It also improves accessibility to public spaces and buildings, making it easier to access services and to engage in community activities. Ultimately, the following [5] Studies can significantly enhance the quality of life for those with visual challenges by boosting their freedom and movement.

We have divided assistive devices that assist visually impaired people for a better understanding of Fig. 1. Assistive devices have evolved on par with technologies and have become increasingly advanced. Physical devices such as white cans and guide dogs have been helping visually impaired people for a long time. As research progresses, various sensor-based systems use built-in sensors and GPS of smartphones to provide location information and directions. The. It has been used to develop various location information and directions, but these can be limited by the battery life of the smartphone, and they may not work indoors.

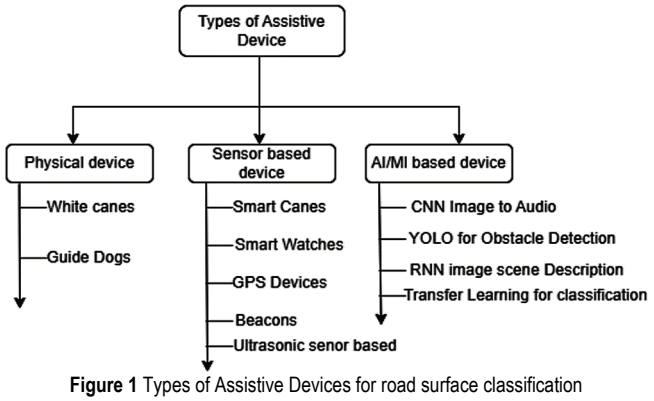

Figure 1 Types of Assistive Devices for road surface classification

Enhancing mobility and independence of visually impaired individuals through artificial intelligence-based road surface detection is an important and significant research area because it aims to improve the lives of visually impaired individuals by utilizing deep learning techniques to detect and classify road surfaces. This technology can help visually impaired individuals navigate more safely and efficiently by providing them with real-time information about road surfaces, such as the presence of obstacles or changes in elevation. Additionally, it can improve the accessibility of public spaces and buildings for visually impaired individuals, making it easier for them to access services and participate in community activities. Overall, this research has the ability to have a positive impact on the daily lives of visually impaired individuals by improving their mobility and independence.

This paper makes the following contributions to assisting visually impaired people.

- Develop a deep-learning-based road surface detection system to assist in the navigation and independence of visually impaired individuals.

- Enhances the precision of detecting road surfaces, diminishing accident risks, and bolstering safety for those with visual impairments.
- Provides an immediate and adaptable method for identifying road conditions, assisting those with visual challenges to traverse unknown terrains more comfortably and assuredly.

The objective of this research is to leverage the capabilities of a pre-trained deep learning model equipped with cutting-edge techniques to detect and classify various road surfaces, in order to aid visually impaired individuals in navigating their surroundings. The author has further enhanced the model's accuracy by incorporating a novel attention mechanism. This system can be seamlessly integrated into mobile devices, such as smartphones, canes, or other assistive devices.

This research will likely involve both theoretical and practical components, such as developing and training deep-learning models, collecting and labelling data, and evaluating the performance of the system in real-world scenarios. Additionally, this research might further examine how the suggested system influences the freedom and movement of those with visual impairments.

The following part reviews previous research on this topic, including deep learning and image recognition. The proposed model is introduced in the Methodology section. The study then addresses the model setting, provides the data information, performs the trials, and compares the outcomes. Finally, the paper concludes with a conclusion.

2 LITERATURE REVIEW

Several authors have emphasized the significance of assistive devices for individuals with visual impairment. In a subsequent paper, the author expands on this topic by conducting four focus groups with assistive technology computer users who are blind or visually impaired to gain broader insights [6]. The objective is to better comprehend how these individuals obtain information about assistive devices and to identify the specific types of information they may possess. The author presents two experiments on Social Interaction Assistants, one of which aims to reduce stereotypical body mannerisms that impede social interactions, while the other is designed to provide individuals with assistive technology to interpret the facial expressions of those they interact with [7].

Advances in CNNs have assisted visually impaired people by enhancing the accuracy and applicability of computer vision systems that are designed to assist them. In recent years, CNNs have been widely used for object recognition [8], visual place recognition [9], verification of CCTV image data through unsupervised learning [10], text classification based on neural networks [11, 12], ANN for estimating cutting forces during helical end milling of metal materials deposited by laser [13], detecting coins and banknotes, and many other applications [14].

A few of these integrate navigation and recognition capabilities into their systems. Based on the above requirements, an assistive device is presented that achieves both capabilities to aid the Visually Impaired person to navigate safely from his/her current location (pose) to a desired destination in an unknown environment and to recognize their surrounding objects. The author described a wearable device designed to help visually impaired individuals navigate unfamiliar environments [15]. The device takes the shape of a pair of eyeglasses, and can help users move safely and efficiently. Additionally, it can help interpret complex surroundings and automatically provide directions on how to move. This study aimed to create a new system that employs OCR and machine learning to assist individuals with visual impairments [16]. Specifically, it develops an indoor item identification system that utilizes a framework based on deep convolutional neural networks. Our objective was to create a robust and reliable solution that can provide visually impaired individuals with an enhanced perception of their surroundings [17].

A new streamlined Convolutional Neural Network (CNN) design was created for the swift recognition of Indian currency notes on web and mobile platforms [18]. The author proposed a walking stick design to help the visually impaired commute to their livelihood [19]. Numerous methods are available to aid blind individuals in navigating their surroundings, including technologies utilizing radio frequency identification (RFID), GPS, and computer vision modules. In a following paper, the author introduced a method for estimating depth from a solitary image, leveraging a local depth assumption without the need for user input. This solution, aimed at aiding individuals with visual impairments, is tailored exclusively for indoor environments such as homes, offices, and businesses [20]. A new system for NAVI was presented based on visual and range information [21]. The author suggested a system that utilizes smartphones to provide navigation assistance, specifically turn-by-turn guidance, through precise and current localization across vast areas [22].

This passage explores several novel strategies aimed at aiding individuals with visual impairments in navigating indoor environments without assistance [23]. One of these strategies is an ambient navigation system that enables free movement without relying on assistance. Another approach utilizes a classification system that employs a Deep Convolutional Neural Network (DCNN) model to identify indoor objects, and this system can be integrated into mobile devices. Moreover, a wearable assistive device shaped like a pair of glasses was presented, which can enhance the user's perception of their surroundings and provide guidance on the direction of movement. Finally, a new indoor object detector was developed using a deep convolutional neural-network-based framework.

Upon analyzing the collected papers, it has been determined that navigation for the visually impaired is a vital area of research, and that deep learning possesses the potential to significantly aid visually impaired individuals in navigating outdoor environments. The results of the literature review indicate that the most recent and pertinent papers

provide invaluable insights that will prove beneficial to our research endeavors. Our thorough examination of these papers has led us to conclude that navigation for visually impaired individuals is an area of utmost importance that warrants further investigation, and that deep learning can play a pivotal role in assisting visually impaired individuals in navigating outdoor environments.

3 PROPOSED METHODOLOGY

This study presents a road-surface detection technique that employs Efficient-Net and a Unified Spatial-Channel attention mechanism [24]. The approach is centered on developing a classification model and incorporating techniques, such as transfer learning and data augmentation, to attain precise automatic categorization of road surfaces. A flow diagram of the proposed method is shown in Fig. 2.

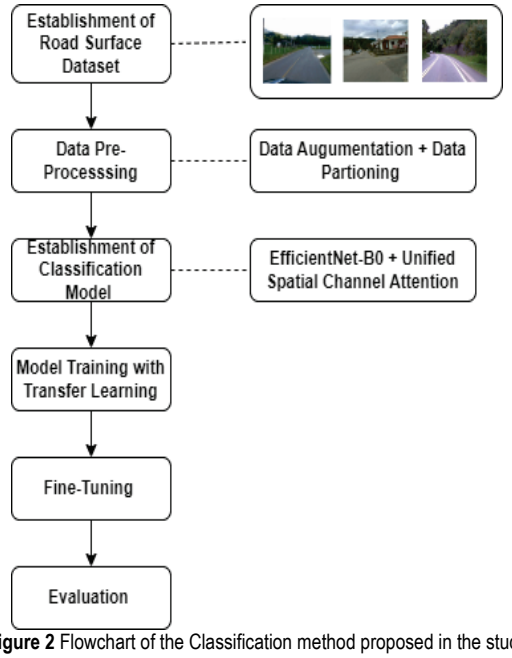

Figure 2 Flowchart of the Classification method proposed in the study

The flowchart in Fig. 2 illustrates a procedure that unites EfficientNet B0, a pre-trained model on ImageNet, with Unified Spatial Channel Attention (USCA) to categorize road surfaces. EfficientNet B0 functions as the foundation for extracting features from raw road-surface images. Subsequently, USCA is integrated, comprising two primary components: Spatial Attention, which concentrates on specific spatial regions, and Channel Attention, which accentuates vital channels. This mechanism improves the feature representation by examining spatial and channel dependencies, resulting in a more refined and focused feature map. The refined features undergo either fine-tuning or transfer learning depending on the chosen strategy. In transfer learning, the knowledge of the pre-trained model is adapted to the road surface dataset, whereas in fine-tuning, the model parameters are adjusted to further refine the pre-trained features for better performance on the road surface images.

Finally, the refined features were classified using a set of fully connected layers, which provided predictions for diverse types of road surfaces. The combination of EfficientNet-B0, USCA, transfer learning, and fine-tuning creates a robust and accurate pipeline for road-surface classification.

3.1 Dataset

We acquired images from the RTK dataset, which were captured using low-cost cameras such as the HP Webcam HD-4110, under real-world conditions [25]. The dataset comprised 77,547 frames from various conditions, including asphalt roads, unpaved roads, and paved roads. From the RTK dataset, we created a dataset consisting of 5,558 images and classified all images into seven different classes. The three classes are asphalt, paved, and unpaved. The dataset was divided into Training, Validation, and Testing sets as shown in Tab. 1. Approximately 70% of the data were in the training folder (4015), 20% were in the testing folder (986), and 10% were in the validation folder (557). The RTK dataset contains real-world images of complex environmental scenarios, such as roads with different vehicles, potholes, and road damage, as shown in Fig. 3.

Figure 3 Sample images from dataset for road surface detection

All images were collected during the daytime with a variety of brightness, texture, and other features. In each roadcategory, there is a slight difference in the surface patterns, such as paved roads that are lighter in color and asphalt roads that are darker in color. We have considered that asphalt roads are roads that do not have any sort of bumps, potholes, or other damage, such as highways and expressways. Unpaved roads are considered bad to walk on because they are not madeup of hard smooth surfaces and have different types of road anomalies. These roads are full of dirt, which is composed of native material on the land

surface. Paved roads are composed of concrete blocks or interlocking. They had different types of patterns on their surfaces. Most pedestrian ways are paved or concrete. We did not perform any type of cropping because we did not want to put extra overhead on computation, which can lead to difficulty in deploying the model in real-time usage.

Table 1 Summary of dataset for road surface detection

Name of class	Train	Test	Valid
Asphalt	1417	343	197
Paved	1386	359	204
Unpaved	1212	284	156

We pre-processed the images before feeding them into the network. We augmented the images to prevent overfitting. Various augmentation techniques, such as geometric transformation, color and contrast adjustments, noise addition, crop, and resize are used, as mentioned in Tab. 2.

Table 2 Different Data Augumentation Techniques Applied

Type	Details
Random Flip	Horizontal and Vertical
Random Rotation	–0.2, 0.2
Random Zoom	0.2
Random Contrast	0.2
Random Translation	0.2, 0.2
Random Height	–0.2,0.2
Random Width	–0.2, 0.2

The implementation of augmentation techniques will enhance the intricacy of our dataset, thereby enabling our model to exhibit more effective generalization capabilities with respect to unseen data. These techniques produce additional variations within an existing dataset without altering the total number of images.

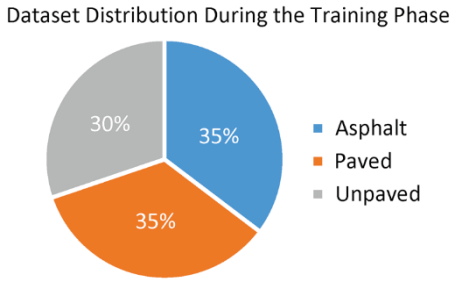

Figure 4 Distribution of dataset for road surface detection for the visually impaired individual

The above pie chart, as shown in Fig. 4, ensures that our dataset is balanced. Balanced dataset analysis has many benefits, such as better generalization to the unseen dataset than to unbalanced data, reduced overfitting, improved model performance, and faster convergence.

3.2 Efficient Net Neural Network

Many pre-trained models are available, but each one is used according to the specific problem domain. The author deals with navigation for visually impaired people, which requires the model to have a smaller size and fewer parameters, making it suitable for a low-edge embedded device such as a smartphone for real-time navigation. The proposed model is based on the EfficientNetB0 architecture [26]. The EfficientNetB0 architecture is a well-known and extensively used network architecture designed for computer vision applications. It is lightweight and can be deployed effortlessly on embedded devices, making it a popular choice in many applications.

Table 3 Size and parameter of different models

Model	No. of Parameters (million)	Size (MB)
EfficientNet-B0 [26]	5.3	350
ResNet-50 [27]	25.6	100
Vgg-16 [28]	138	553
DenseNet-121 [29]	8.8	100

From Tab. 3 EfficientNet-B0 can act as a suitable model for assisting visually impaired people because of its size and the number of parameters used, which makes it suitable for real-time operations. Despite being lightweight, EfficientNet-B0 is known for its good performance in a variety of computer vision tasks, including object detection and image classification. EfficientNet-B0 can be fine-tuned for a specific task using a smaller dataset. It has a faster inference time owing to its small size, which is important for real-time applications. Another advantage of EfficientNet-B0 is its adaptive architecture, which allows it to easily adapt to different input sizes and resolutions.

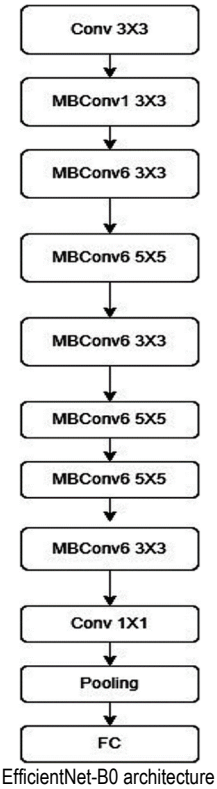

Figure 5 EfficientNet-B0 architecture flowchart

EfficientNet-B0, the foundational model of the EfficientNet family, offers a novel and holistic approach to neural network architecture optimization, balancing both accuracy and computational efficiency. The architecture of

EfficientNet-B0 is shown in Fig. 5. At the heart of its design is the innovative concept of compound scaling, a method that diverges from traditional practices by uniformly scaling the depth, width, and resolution of the network, as defined in Eq. (1). This technique ensures that no single dimension is overoptimized at the expense of the others. As the architecture delves deeper, these blocks, equipped with squeeze-and-excitation operations, manage the intricate task of learning channel-wise dependencies, thereby ensuring a comprehensive understanding of features. Beginning with a convolutional stem that transforms the 3-channel RGB input into a 32-channel feature map, the network sequences through a series of MBConv blocks.

The core architecture involved MBConv blocks equipped with squeeze-and-excitation operations. These blocks manage channel-wise features to understand intricate details in the images. Starting with a convolutional stem that processes RGB inputs into a feature map, the network navigates through these blocks to adaptively recalibrate features, making the model more perceptive of essential information.

EfficientNet-B0 boasts a unique design along with practical techniques such as DropConnect regularization to prevent overfitting during training. This model utilizes global average pooling to compress spatial dimensions and a fully connected layer for final classification while maintaining accuracy across varying computational budgets. Its goal is to provide improved accuracy and faster inference without sacrificing model size or complexity. The inclusion of compound scaling and efficient channel attention allows better performance and adaptability in various scenarios.

$$d = \alpha^{\phi}, w = \beta^{\phi}, r = \gamma^{\phi}. \tag{1}$$

Where EffiecientNetB0 introduced scaling in the depth d, width w, and resolution r. α, β, and γ are scaling coefficients, and ϕ symbolizes the scaling factor that controls the extent to which the depth, width, and resolution of the network should be adjusted.

The essence of Eq. (1) allows EfficientNet to adjust its model complexity effectively by manipulating the depth, width, and resolution through the application of scaling coefficients and a scaling factor. This approach enables the model architecture to be tailored for different computational budgets while striving to preserve high accuracy.

$$F_{\mathrm{sq}}(u_{\mathrm{c}}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_{\mathrm{c}}(i,j). \tag{2}$$

The following equation Eq. (2) calculates the average value of all the elements in the tensor by adding all its elements and then dividing this sum by the total number of elements, which is $H \times W$. This process compresses the spatial information and reduces the spatial dimensions of the tensor while preserving important information for subsequent operations in the network.

$$S = \sigma\left(\boldsymbol{W}_2 \partial(\boldsymbol{W}_1 z)\right), \tag{3}$$

where S represents the output of the excitation operation, $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ represent the weight matrices; z is the output of the squeeze block, $g(z, W)$ represents an intermediate computation; ∂ refers to the intermediate computations including ReLU operation and where σ is the sigmoid function.

The above equation Eq. (3) is vital in channel attention mechanisms because it enables the neural network to concentrate on critical channels by dynamically adjusting the significance of each channel in the feature map (z). Weight matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are learned during training to emphasize the relevant channels while deemphasizing the less informative ones. The excitation operation facilitates the network's ability to efficiently capture channel-wise dependencies, leading to enhanced feature representation and improved global information access for superior decision making.

3.3 Unified Spatial-Channel Attention

Attention mechanisms play a crucial role in assisting neural networks to concentrate on essential input data, thereby enhancing their learning capabilities and predictive accuracy. They are particularly advantageous in handling variable sequence lengths because they enable the network to focus adaptively on different input segmment.

Attention mechanisms contribute to an improved model performance by capturing intricate patterns and long-range dependencies. Moreover, the transparency provided by attention mechanisms helps clarify the significance of input elements in the decision-making process. Attention mechanisms optimize the computational efficiency and processing speed by directing attention to specific elements.

Spatial attention focuses on spatial relationships within an image, as defined in Eq. (4), by focusing on specific regions or pixels relevant to the task. It helps models highlight critical spatial features such as edges or textures, enabling them to identify key visual patterns.

$$S = \sigma\left(\boldsymbol{W}_2 \delta(\boldsymbol{W}_1 z)\right), \tag{4}$$

where $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ represent the weight matrices; δ denotes ReLU operation; where σ is the sigmoid function and z is the output of the spatial squeeze block.

By contrast, channel attention operates across channels or feature maps, as defined in Eq. (5), which allows the model to assign different weights to each channel based on its importance. By capturing channel-wise dependencies, it refines feature representations and enhances the model's understanding of the semantic information in the data. When combined with unified spatial-channel attention, these attention mechanisms enable the network to discern both spatial and semantic details, optimizing its ability to extract meaningful information from images.

$$CA = \sigma\left[\boldsymbol{W}_2\delta\left(\boldsymbol{W}_1 \cdot F_{\mathrm{sq}}(u_{\mathrm{c}})\right)\right], \quad (5)$$

$F_{\mathrm{sq}}(u_{\mathrm{c}})$ denotes the spatially squeezed representation of the channel.

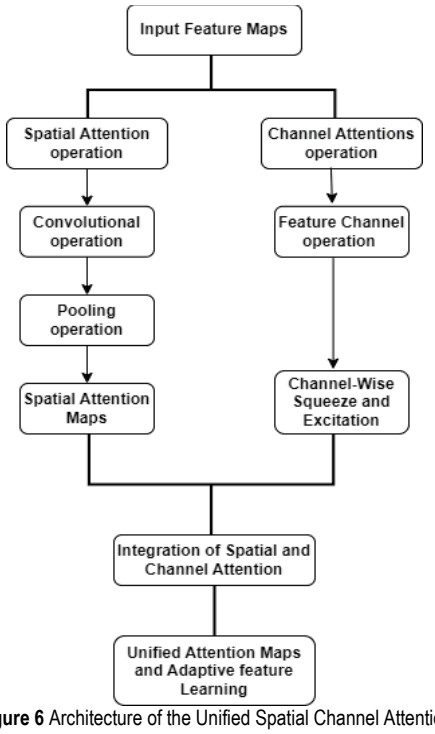

Figure 6 Architecture of the Unified Spatial Channel Attention

The flowchart in Fig. 6 shows the unified spatial-channel attention mechanism, which begins with the input data and proceeds through convolutional operations for feature extraction. The process then branches into spatial and channel attention modules. In the spatial attention module, convolutional layers are followed by global pooling operations, highlighting spatial details. Conversely, the channel attention module utilizes operations such as squeeze-and-excitation, focusing on the feature interdependencies within each channel. These separate pathways converge, allowing the unified attention mechanism to successfully combine the spatial and channel-wise information. The resulting attentional output is reintegrated into the network, enhancing feature representation and aiding classification tasks. Ultimately, this unified spatial channel attention mechanism harmoniously merges spatial and channel-dependent features, improving the model comprehension and classification accuracy, which can increase the accuracy of our road surface classification and generalize better for the task.

3.4 Transfer Learning

The deep learning algorithm addresses the limitations of traditional machine learning methods in feature extraction; however, it requires many images for training to achieve high accuracy. In addition, creating a diverse dataset of rock images can be time consuming. To overcome this issue, transfer learning was used in this study to fine-tune a specific model using the parameters and weights of a pretrained model trained on a large-scale annotated image dataset. By re-training and fine-tuning the specific model, a more accurate classification model can be obtained using fewer rock images and a shorter training time. In road surface classification using transfer learning, the process begins with the selection of a suitable pretrained model for feature extraction. The EfficientNet-B0 architecture was used as the pre-trained model, and customization involved loading the pre-trained EfficientNet-B0 weights and freezing most of the layers to preserve the learned features while fine-tuning only the final layers for the specific task. This methodology captures generalized features from broader image datasets and refines them to cater to the nuances of road-surface classification. By freezing the layers, the model optimizes computational efficiency and reduces the need for extensive training on the new dataset.

After this adaptation, the model proceeds to a training phase with the road surface dataset, allowing it to learn task-specific features while benefiting from the generalizable knowledge initially obtained from the pretrained model. Through this sequential process, the model gained insights into the distinctive characteristics of road surfaces, leveraging the foundational knowledge acquired from its pre-trained state to enhance its classification capabilities.

3.5 Fine Tuning

Fine-tuning is a transfer learning technique that entails further training of a pre-trained model on a new dataset while retaining the knowledge it previously acquired. This approach builds upon the weights learned during the initial training and adjusts them to suit the new task or dataset better.

By fine-tuning the pre-trained EfficientNet-B0 model, which has already gained knowledge about image features and patterns through its previous training, we aimed to refine the model's ability to recognize road surface features. We achieve this by making slight adjustments to the learned features such that they align better with the unique features of our new dataset.

Table 4 Hyperparameters used in road surface classification

Parameters	Value
Optimizer	Adam, RMSprop
Learning rate	1×10^{-5}
Batch Size	10
Dropout	0.3 to 0.5
Early Stopping	Validation loss, Patience = 5
Activation Function	SoftMax function
Loss function	Categorical Cross Entropy

The fine-tuning process involved two strategic steps. In the first step, the core layers of the model remain unaltered, whereas we focus on optimizing the newly integrated classification components, such as global average pooling and dense layers, all hyperparameters details are mentioned in Tab. 4. This phase is critical for adapting the model to discern unique road surface attributes identified during the

transfer learning stage. The second step involves refining the accuracy of the model in road surface classification by fine-tuning specific advanced sections without altering the foundational layers. Utilizing the RMSprop optimizer, these adjustments aim to amplify the model's discernment of crucial spatial nuances that are essential for accurate classification.

Overall, this methodical fine-tuning approach meticulously tailors the EfficientNet-B0 architecture, bolstered by unified spatial channel attention, to excel in discerning the intricate features inherent in road surface images.

4 RESULT AND DISCUSSION

We tested the model for detecting road classification for visually impaired pedestrians and compared it with the basic model EfficientNet-B0 [26] and traditional machine learning algorithms, such as ResNet50 [27] and Random Forest [30], which are shown in Tab. 5. Compared to the models mentioned above, our approach provides promising results. We conducted this experiment using our hand-labelled dataset, which includes three distinct categories: Asphalt, Paved, and Unpaved. We allocated 70% of the data for training, 20% for testing, and the remaining 10% for validation. We calculated the F1-score for each class as false negatives and false positives, which are more important than true negatives and true positives, as in our case, the dataset was not balanced. We ran this test on our manually classified dataset which contains 4015 training images, 557 validation images, and 986 testing images. In the dataset, we included real-world images, including images with other vehicles, while avoiding images that contained transitions between road surfaces and frames that consist of the very strong glare of sun rays causing reflection. Even after including images with complex conditions, our approach can detect the surfaces of vehicles with good accuracy. The Confusion matrix helps us understand the model performance for all classes of the dataset. The matrix compares the actual target with those predicted by our road surface quality classification model.

In both machine learning and statistics, the confusion matrix is a crucial instrument for assessing the performance of the classification models. It offers a detailed comparison of predicted results against true values, shedding light on the model's overall precision and the nature of mistakes it commits. Fundamentally, in tasks involving multiclass classification, there are four primary components: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), where TP and TN capture accurate predictions, and FP and FN highlight instances in which the model's prediction contradicts the actual outcome. By diving deeply into these elements, we can pinpoint areas where the model falls short, underscoring the essential role of the confusion matrix in fine-tuning and enhancing classifiers.

The model demonstrates as in Fig. 7 strong discernment between Asphalt and Unpaved classes, with high precision and true positive rates. However, it shows a slightly higher tendency for misclassification within the Paved class, resulting in a few false positives and false negatives. Despite this, the model maintains a high overall accuracy, especially in distinguishing between Asphalt and Unpaved surfaces, and has a lower rate of misclassifications within the Paved class.

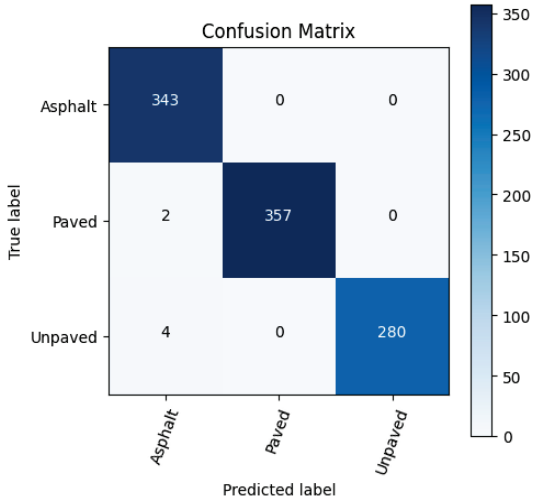

Figure 7 Confusion Matrix for road surface classification

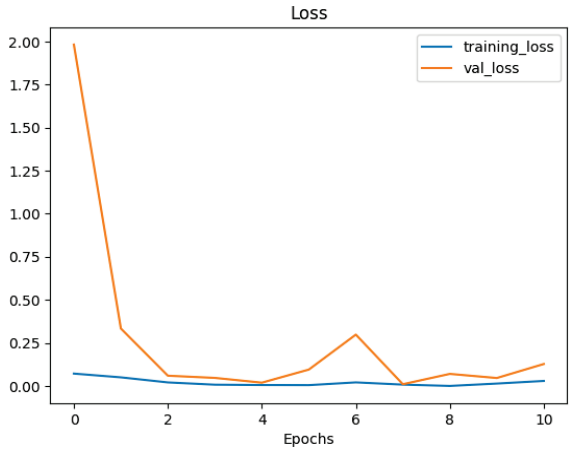

Figure 8 Training loss and validation loss graph for the proposed model

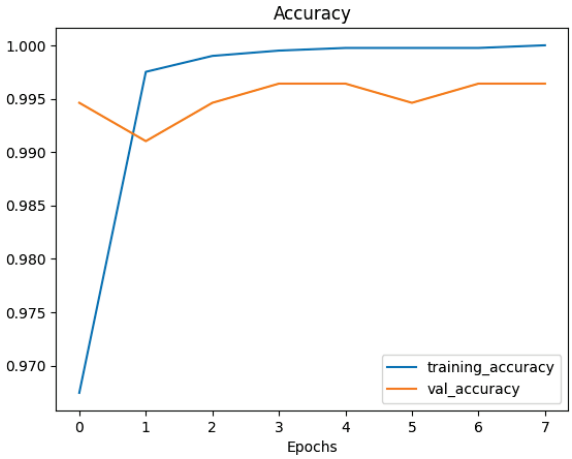

Figure 9 Training accuracy and validation accuracy graph for the proposed model

Fig. 8 and Fig. 9 show the accuracy and loss values during the training and validation phases of our modified EfficientNetB0, respectively. The proposed novel

architecture, based on EfficientNetB0 with the fusion of a unified spatial channel attention mechanism, achieved 99.39% testing accuracy, whereas EfficientNetB0 achieved 96.85% accuracy. The modelis deployed in the form of a web application on the Heroku platform using the Flask application, which provides satisfactory resultsin real-world conditions.

The data presented in Tab. 5 unequivocally demonstrates that the proposed model outperformed the other models by a considerable margin.

Table 5 Comparative analysis of accuracy of different models

Classes	EfficientNet [26] (%)	ResNet [27] (%)	Random Forest [30] (%)	Proposed Model (%)
Asphalt	97.80	92.56	72.34	99.24
Paved	96.23	93.34	82.67	99.68
Unpaved	96.54	89.12	78.89	99.15

Table 6 Classification report of proposed model

Name of the class	Precision (%)	Recall (%)	F1-score (%)	Accuracy
Asphalt	98.28	100	99.12	99.24
Paved	99.44	99.44	99.44	99.68
Unpaved	98.59	98.59	98.59	99.15

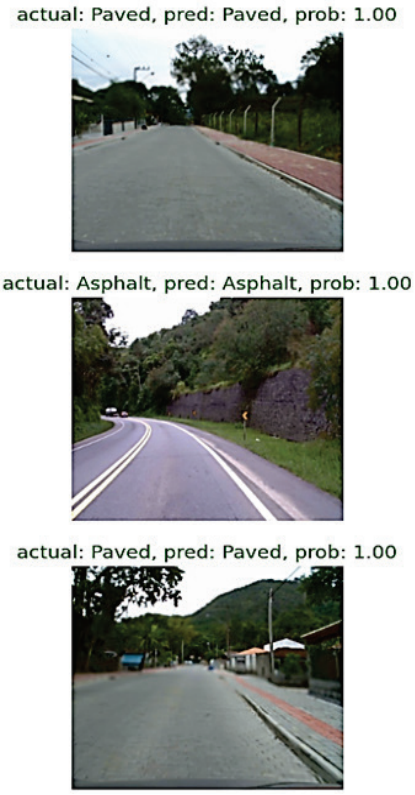

Figure 10 Output of the proposed model based on EfficientNet-B0 and Unified spatial channel attention

A classification report serves as an instrument for machine learning to assess the performance of classification models. It provides an overall assessment of the accuracy of the model as well as class-specific evaluation, which helps identify which classes require improvement. Tab. 6 shows the classification report by which we can further determine the performance of the proposed model over individual classes.

The output of the proposed model is shown in Fig. 10, and it was clearly able to predict the road surface with good accuracy.

5 CONCLUSION AND FUTURE WORK

Previous research on visually impaired pedestrians has largely focused on the detection of obstacles in their paths to help them avoid potential hazards. However, our work has focused on assessing the quality of the road surface upon which these individuals must navigate. By increasing awareness of their surroundings, this approach can aid visually impaired pedestrians in adjusting their walking patterns and speeds. In this study, we present a novel architecture based on EfficientNetB0 with a unified spatial channel attention mechanism that achieves state-of-the-art results and outperforms both individual models and traditional machine-learning algorithms. Our proposed model achieved an accuracy of 99.39%, surpassing the 96.85% achieved by EfficientNetB0. Additionally, our model is well suited for deployment on embedded devices with limited computational power. Experimental results confirm the efficacy of our proposed approach, and we plan to further expand our research by incorporating additional classes while maintaining high accuracy and by identifying various obstacles and potential hazards on the road surface, including stray animals.

6 REFERENCES

[1] World Health Organisation. (2020). Visual impairment and blindness. Retrieved from https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment.

[2] Helal, A., Moore, S. E. & Ramachandran, B. (2001). Drishti: an integrated navigation system for visually impaired and disabled individuals In *Proceedings of the Fifth International Symposium on Wearable Computers*, 149-156. https://doi.org/10.1109/ISWC.2001.962119

[3] Ramadhan, A. J. (2018). Wearable Smart Systems for Visually Impaired People: *Sensors, 18*(3), 843. https://doi.org/10.3390/s18030843

[4] Porzi, L., Messelodi, S., Modena, C. M. & Ricci, E. (2013). Smart watch-based gesture recognition system for assisting people with visual impairments. In *Proceedings of the 3rd ACM International Workshop on Interactive Multimedia on Mobile and Portable devices*, 19-24. https://doi.org/10.1145/2505483.2505487

[5] Gerber, E. (2003). The Benefits of and Barriers to Computer use for visually impaired individuals *Journal of Visual Impairment & Blindness, 97*(9), 536-550. https://doi.org/10.1177/0145482X0309700905

[6] Krishna, S. & Panchanathan, S. (2010). Assistive technologies as effective mediators in interpersonal social interactions for persons with visual disabilities In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds) *Computers Helping People with Special Needs, ICCHP 2010, Lecture Notes in Computer Science, vol 6180*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-14100-3_47

[7] Aziz, N., Roseli, N. & Mutalib, A. A. (2011). Visually Impaired Children's acceptance of assistive courseware. *American Journal of Applied Sciences, 8*, 1019-1026. https://doi.org/10.3844/ajassp.2011.1019.1026

[8] Shah, S., Bandariya, J., Jain, G., Ghevariya, M. & Dastoor, S. (2019). CNN based Auto-Assistance system as a boon for directing visually impaired persons. In *The 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 235-240. https://doi.org/10.1109/ICOEI.2019.8862699

[9] Fang, Y., Wang, K., Cheng, R., Yang, K. & Bai, J. (2019). Visual place recognition based on multilevel descriptors for visually impaired people. *Proc. SPIE 11158, and Target and Background Signatures V* 1115808. https://doi.org/10.1117/12.2532524

[10] Lee, Y. (2023). Verification of CCTV image data through unsupervised learning model of deep learning. *Tehnički glasnik, 17*(3), 353-358. https://doi.org/10.31803/tg-20221227094126

[11] Kim, D. (2023). Text classification based on neural-network fusion. *Tehnički glasnik, 17*(3), 359-366. https://doi.org/10.31803/tg-20221228154330

[12] Lee, S. (2023). Text classification of mixed models based on deep learning. *Tehnički glasnik, 17*(3), 367-374. https://doi.org/10.31803/tg-20221228180808

[13] Župerl, U. & Kovačič, M. (2023). Artificial Neural Network System for Predicting Cutting Forces in Helical-End Milling of Laser-Deposited Metal Materials. *Tehnički glasnik, 17*(2), 223-230. https://doi.org/10.31803/tg-20230417145110

[14] Alghamdi, S. (2019). Shopping and tourism for blind people using RFID as an IoT application. In *The 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, 1-4. https://doi.org/10.1109/CAIS.2019.8769581

[15] Bai, J., Liu, Z., Lin, Y., Li, Y., Lian, S. & Liu, D. (2019). Wearable travel aids in the environment perception and navigation of visually impaired people. *Electronics, 8*(6), 697. https://doi.org/10.3390/electronics8060697

[16] Kaur, B. & Bhattacharya, J. (2019). A scene perception system for the visually impaired based on object detection and classification using a multimodal DCNN. *Journal of Electronic Imaging, 28*(01), 1. https://doi.org/10.1117/1.JEI.28.1.013031

[17] Afif, M., Ayachi, R., Said, Y., Pissaloux, E. & Atri, M. (2020). An evaluation of RetinaNet for indoor object detection for blind and visually impaired persons assists navigation. *Neural Processing Letters, 51*(3), 2265-2279. https://doi.org/10.1007/s11063-020-10197-9

[18] Veeramsetty, V., Singal, G. & Badal, T. (2020). Coinnet: A platform-independent application to recognize Indian currency notes using deep learning techniques. *Multimed Tools Appl, 79*, 22569-22594. https://doi.org/10.1007/s11042-020-09031-0

[19] Kanna, S. B., Kumar, T. G., Niranjan, C., Prashanth, S., Gini, J. R. & Harikumar, M. E. (2021). Low-Cost Smart Navigation System for Blinds. In *The 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 1*, 466-471. https://doi.org/10.1109/ICACCS51430.2021.9442056

[20] Praveen, R. G. & Paily, R. P. (2013). Blind Navigation Assistance for Visually Impaired based on the Local Depth Hypothesis from a Single Image. *International Conference on Design and Manufacturing (IConDM2013), Procedia Engineering, 64*, 351-360. https://doi.org/10.1016/j.proeng.2013.09.107

[21] Aladrén, A., López-Nicolás, G., Puig, L. & Guerrero, J. J. (2016). Navigation Assistance for the Visually Impaired Using RGB-D Sensor with Range Expansion. *IEEE Systems Journal, 10*(3), 922-932. https://doi.org/10.1109/JSYST.2014.2320639

[22] Ahmetovic, D., Gleason, C., Kitani, K. M., Takagi, H. & Asakawa, C. (2016). NavCog: A turn-by-turn smartphone navigation assistant for people with visual impairment or blindness. In *Proceedings of the 13th International Web for All Conference*, 1-2. https://doi.org/10.1145/2899475.2899509

[23] Chaccour, K. & Badr, G. (2016). Computer vision guidance system for indoor navigation of visually impaired people. At *The 8th IEEE International Conference on Intelligent Systems (IS)*, 449-454. https://doi.org/10.1109/IS.2016.7737460.

[24] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N. & Polosukhin, I. (2017). Therefore, attention is required. *Advances in neural information processing systems*, *30*.

[25] Rateke, T., Justen, K. A. & von Wangenheim, A. (2019). Road surface classification with images captured from low-cost camera-road traversing knowledge (RTK) dataset. *Revista De Informática Teórica E Aplicada, 26*(3), 50-64. https://doi.org/10.22456/2175-2745.91522

[26] Tan, M. & Le, Q. (2019). EfficientNet: Rethinking model scaling for Convolutional Neural Networks. *Proceedings of the 36th International Conference on Machine Learning, 97*, 6105-6114. Retrieved from https://proceedings.mlr.press/v97/tan19a.html.

[27] 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 (CVPR2016)*, Las Vegas, NV, USA, 770-778. https://doi.org/10.1109/CVPR.2016.90

[28] Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[29] Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700-4708.

[30] Louppe, G. (2014). Understanding random forests: From theory to practice. *arXiv preprint arXiv:1407.7502*. https://doi.org/10.48550/arXiv.1407.7502

Author's contacts:

Amit Chaudhary, Research Scholar
(Corresponding Author)
Harcourt Butler Technical University,
Nawabganj, Kanpur, Uttar Pradesh 208002, India
amitchaudhary.gkg@gmail.com

Prabhat Verma, Associate Professor Dr.
Harcourt Butler Technical University,
Nawabganj, Kanpur, Uttar Pradesh 208002, India
pverma@hbtu.ac.in