Enhancing Outdoor Mobility and Environment Perception for Visually Impaired Individuals Through a Customized CNN-based System

Athulya N K[1], Sivakumar Ramachandran[2], Neetha George[3], Ambily N[4], Linu Shine[5]*
Department of Electronics and Communication, College of Engineering Trivandrum, India[1]
Department of Electronics and Communication, Government Engineering College Wayanad, India[2]
Department of Electronics and Communication, Rajeev Gandhi Institute of Technology Kottayam, India[3,5]
Department of Electronics and Communication, Government Engineering College Idukki, India[4]

***Abstract*—Visual impairment indicates any kind of vision loss including blindness. Individuals with visual impairments face significant challenges when trying to perceive their surroundings from a global perspective and navigating unfamiliar environments. Existing assistive technologies predominantly focus on obstacle avoidance, neglecting to provide comprehensive information about the overall environment. To address this gap, the proposed system employs a customized Convolutional Neural Network (CNN) model tailored to accurately predict the type of outdoor ground terrain the user is traversing. This information is then conveyed to the user audibly. It can also detect the presence of puddles on the road and let the user know whether the outside floor is wet (slippery). The proposed deep-learning architecture is trained on images collected from sources including the Stagnant Water dataset, the GTOS-Mobile dataset and a custom dataset. The trained model is then integrated into an Android app, providing visually impaired (VI) people with effective surrounding perception capabilities, leading to better travel and, ultimately, better living.**

***Keywords*—Visually impaired; terrain identification; puddle detection; deep learning**

I. Introduction

Visual impairment encompasses a range of conditions that can result in partial or complete loss of vision, including color blindness, affecting over 250 million people worldwide. Such individuals encounter difficulties while comprehending and interacting with their environment due to their limited visual contact with their surroundings. Physically moving around can be particularly challenging for them as they struggle to identify their location and navigate to different places.

For navigation, individuals with visual impairments often utilize white canes or guide dogs, although the latter option can be costly. However, despite the assistance provided by white canes, the task of maneuvering through unfamiliar environments continues to pose a significant challenge. Conventional navigation techniques primarily focus on obstacle avoidance [1], which curtails the capacity of visually impaired (VI) individuals to engage with novel surroundings fully. Consequently, many still face obstacles in gaining a holistic perception of their environment.

Recent advancements in technology have paved the way for the development of intelligent or augmented white canes that utilize image processing and deep learning techniques to aid individuals with low vision in obstacle avoidance, object detection [2], and indoor/outdoor navigation [3]. However, these systems have a limitation in that they do not provide a comprehensive perception of the surrounding environment. For instance, while these canes assist in identifying objects and navigating outdoors, they cannot provide information on the terrain, which is crucial for individuals with visual impairments to determine the floor they are walking on. By incorporating terrain information into these systems, VI individuals can benefit from a more comprehensive and seamless travel experience, ultimately leading to an improved quality of life. This is particularly important for elderly individuals with low vision.

The proposed method makes significant contributions to assistive technology for VI individuals, with a strong focus on human needs and safety. The main contributions are:

- Proposed a novel custom CNN for Comprehensive Terrain Identification: An innovative lightweight CNN has been created to recognize diverse terrains. Through the integration of this CNN, an Android application is developed for identifying different types of terrain. This pioneering functionality effectively addresses a significant deficiency in current systems, delivering vital insights to VI users regarding the specific characteristics of the ground they are traversing.
- Hazard Detection: The system's ability to detect road puddles or wet floors is a key contribution that sets it apart from conventional intelligent white cane systems. By proactively alerting users to potential hazards, the approach reduces the risk of slips, falls, and accidents, ensuring a safer travel experience for individuals with low vision, especially the elderly.
- Developed a dataset to train and test the proposed model: A dataset is made by combining standard datasets like the Stagnant water dataset, and the GTOS-Mobile dataset, collecting images from Google and images captured using a mobile phone.
- Real-time Audio Feedback: To promote effective communication between the system and the user, the integration of audio feedback proves invaluable. By conveying information through auditory cues, the system ensures that VI individuals receive timely updates

*Corresponding authors.

about the terrain and detected hazards, allowing them to make informed decisions promptly.

In summary, our approach introduces a novel lightweight CNN with the ability to identify different terrains. We curate a specialized dataset for training and model development. Employing this innovative CNN, an Android application is developed that offers terrain insights and alerts users through audio feedback, particularly notifying them about possible hazards such as slippery floors or puddles. This approach prioritizes human needs, safety, and independence, with the potential to revolutionize assistive technology and significantly enhance the lives of individuals with limited vision, particularly elderly users, leading to an overall improved quality of life.

The structure of the paper is as follows: In Section II, we delve into the most recent advancements in intelligent tools designed[9] proposed a system explicitly for individuals with visual impairments. Section III provides a comprehensive explanation of our proposed methodology. We outline the datasets utilized in this study in Section IV. The experimental validation and outcomes of our proposed system are detailed in Section V. Finally, Section VI[9] proposed a system serves as the conclusion of this paper.

II. Related Works

Vision forms one of the most important sense organs, which gives vital information about surroundings. Our sense of sight is responsible for 80 percent of what we perceive. Vision enables humans to interact with their surroundings. One of the side effects of vision loss is a lack of confidence in one's ability to travel safely. The placement of tactile ground surface indicators, unsafe sidewalks, and the presence of barriers on sidewalks are key challenges in comprehending the outdoor difficulties experienced by VI individuals [4]. This section reviews some of the critical works in the field of guiding aids for VI people. This analysis is split into related works regarding existing smart white canes and terrain recognition.

A. Robotic White Cane

The white cane is the most frequently used aid by individuals with visual impairments. Its main function is to help users assess their environment for possible dangers. Additionally, it aids in signaling to others that the user is blind, ensuring they receive appropriate assistance. In recent years, numerous enhanced versions of the white cane have emerged, offering a range of capabilities. Many of these innovations concentrate on tasks such as avoiding obstacles, identifying objects, and facilitating outdoor navigation.

Anwar et al. [1], introduces a smart cane equipped with an alarm to aid individuals with visual impairments when navigating challenging paths. An RF remote transmitter and receiver make it a unique electronic stick. It is equipped with an ultrasonic sensor and a buzzer. The VI person's walking path is detected using an ultrasonic sensor. At the same time, a global positioning system (GPS) paired with a voice stick for navigation, allows users to learn their present location and distance from their destination via voice commands. However, this system is constrained to the local vicinity only. [5] proposed a smart walker navigation system that assists VI individuals with difficulty walking. It is utilized for two purposes: local obstacle detection in the spatial information setting, and guided navigation for achieving the desired goal. Vibrotactile signals provide obstacle information or navigation commands to the user.

A voice-activated electronic stick whose goal is to give users the confidence to move around in new situations was introduced by [6]. This enhanced model comprises global system for mobile communications (GSM), GPS , and Ultrasonic technology. It also employs biological authentication and incorporates an emergency trigger in the alarm system. [7] proposed an enhanced white cane that aids the blind community in guiding by providing results for all 270 degrees from the smart security walking stick's position. Ultrasonic sensors with a wide beam angle assist in a variety of obstacle detection applications. It performs well at identifying obstructions in the user's path within a three-meter range. This technology is low-cost, lightweight, and energy-efficient, with a noticeable quick response time.

The traditional VI aiding technologies addressed the obstacle avoidance problem alone. Information about the terrain is also vital for safe and smooth navigation for the VI person. A generative adversarial network (GAN) model named DiscoGAN is created by [8] to effectively translate ground images into a tactile signal that can be presented by an off-the-shelf vibration device. In this way, it can provide a global perception of the surroundings. The research [9] proposed a system based on DeepLabv3+ that is an improved semantic segmentation network. The technology can be used on a mobile phone to assist VI people in both indoor and outdoor settings. The training dataset is preprocessed with an illumination-invariant transformation to reduce the impact of variations in light. Bashiri et al. introduced a novel framework in their study [10], employing deep neural networks to facilitate indoor navigation for visually impaired individuals.

Existing literature for puddle detection mainly uses object detection models. Some of the existing smart white canes include puddle or water detection using a moisture sensor. But if the battery is not charged, this will not work which is the main disadvantage faced by these augmented canes [11].

B. Ground Terrain Recognition

The quality of the terrain dataset, the feature extraction method, and the classification algorithm for terrain features are the key factors that affect the performance of a terrain image recognition system [12]. These aspects are complementary, necessitating optimization and control throughout, to improve the terrain classification and generalization capacity of a new algorithm. Moreover, there exists a lack of publicly available terrain classification datasets for research purposes. The literature related to terrain recognition explained here used proprietary datasets in their study.

Terrain identification is critical for outdoor mobile robot gait planning, speed control, and observation of the surroundings, among other things. The study [13] opted for a system in which an outdoor mobile robot is built that runs on a terrain dataset and extracts the high-level elements of the terrain image from MobileNet and DenseNet using the migrating learning approach. Extent-of-Texture information proposed by

[14] is another unique way to ground-terrain recognition, which employs a CNN backbone feature extractor network to extract relevant information from ground terrain images and to model the extent of texture and shape information locally. The research [15] proposed a differential angular imaging Network (DAIN), where small angular variations in image capture provide an enhanced appearance representation that significantly improves recognition. The features of materials recorded in the angular and spatial dimensions are encoded in this innovative network architecture. A Deep Encoding Pooling network [16] is another method for ground terrain categorization. Semantic scene comprehension is critical for gaining a better grasp of the surroundings. The study [17] proposed a system based on a combination of the efficient residual factorized network(ERFNet) and the 3D point cloud with a pyramid scene parsing network (PSPNet). The above-mentioned models are complex models for which accuracy can be further improved. Furthermore, most of them have not been implemented in real time. Recently, Intelligent Vehicles (IV) methodologies have been applied to the development of navigation assistive systems for VI people.

The analysis of existing literature highlights significant research gaps in the field. Most robotic white canes prioritize addressing obstacles, overlooking the potential of offering a comprehensive understanding of the environment, especially in unfamiliar routes. Presently, there's a lack of systems that can effectively notify users about hazards like puddles, wet and slippery floors. Additionally, the accuracy of current ground floor classification systems could be enhanced. The primary objective of our proposed system is to furnish users with real-time information about outdoor ground conditions, particularly alerting them to road puddles and wet floors, through a custom CNN model that accurately predicts terrain. This crucial information is conveyed audibly to enhance user safety and experience.

III. Methodology

The prevailing terrain recognition models predominantly feature intricate deep-learning architectures. Nonetheless, for individuals with visual impairments, conveying terrain details through an Android app holds more promise. This research strives to forge a path towards a simplified and computationally streamlined deep-learning model tailored for terrain recognition systems. To realize this vision, a direct approach is adopted, entailing the deployment of a specialized CNN model. Additionally, an Android application is meticulously crafted to enhance accessibility for VI individuals, facilitating seamless access to crucial terrain information.

A custom dataset consisting of five image classes, namely cement, grass, road, wet floor, and puddles has been created to help VI people better understand their terrain and be alerted of any potential hindrances. This dataset is a combination of GTOS-Mobile (Ground Terrain Outdoor Services) [18] data, Google images, IEEE Stagnant Water dataset, and ground terrain images captured using a mobile phone. An Android app has been developed which uses a custom-built CNN to detect ground terrain and provides audio feedback to the user with directions based on the detected features in real-time.

The proposed pipeline consists of three main steps, namely data pre-processing, CNN model design, and Android app development. The images obtained from various data sets are resized to the same dimension in the pre-processing phase and are then augmented before feeding to the custom CNN. The pre-processing steps used were resizing, and gray scaling. The data augmentation involves random flip, random zoom, and random rotation. After model training, test data prediction and evaluation are done using the trained deep learning model. Fig. 1 shows the flow diagram of the proposed system for ground terrain detection. It represents the schematic of the proposed system.

A. Custom CNN

The Custom CNN architecture is comprised of five convolutional layers as depicted in Fig. 2. Detailed insights into the layer structure and parameter specifics can be found in Table I. The initial convolution layer block consists of two sets of convolution layers, each comprising 16 filters. The subsequent block contains two sets of convolution layers with 32 filters each, followed by a third block with a convolution layer featuring 64 filters. These layers employ the same padding scheme, along with ReLU activation and Max pooling. The filter size utilized in each of these convolutional layers is 3x3.

After these convolutional layers, a flattening layer is introduced, followed by a fully connected layer consisting of 128 filters, employing ReLU activation. Ultimately, a dense layer is incorporated, with the number of classes serving as a parameter. In summary, the Custom CNN model includes 560,000 trainable parameters, featuring five convolutional layers with associated Maxpooling layers and ReLU activation. It further incorporates a fully connected layer utilizing a dropout rate of 50%, followed by an output dense layer with the number of classes as a parameter. The model employs the sparse categorical cross-entropy loss function, particularly useful for multi-class image classification tasks. This loss function quantifies the cross-entropy loss between predictions and labels. Importantly, its utilization of integer representations for labels instead of vector representations contributes to efficiency in memory and computation. The mathematical formulation of this entropy function is:

$$L_{\mathrm{CE}} = -\sum_{i=1}^{n} t_i \log\left(p_i\right)$$

where t_i is the truth label and p_i is the Softmax probability for the i^{th} class and n represents the number of classes.

The architecture of the Custom CNN is depicted in Fig. 2. To make the test results more accurate, we fine-tuned the CNN model so that it could better understand the test data. By doing this, the model became better at recognizing different features in the test images. This process involves making small improvements to how the model learns from the given data. The combination of the CNN's design, how it extracts features, and the fine-tuning process all work together to create a strong model that can accurately identify and classify images. The images were resized to dimensions of 64×64 pixels and a batch size of 15 is selected empirically. The number of filters within the fully connected Layer is fixed as 128.

IV. Data Collection

We curated an exclusive dataset tailored to our proposed system. This dataset comprises diverse visual representations

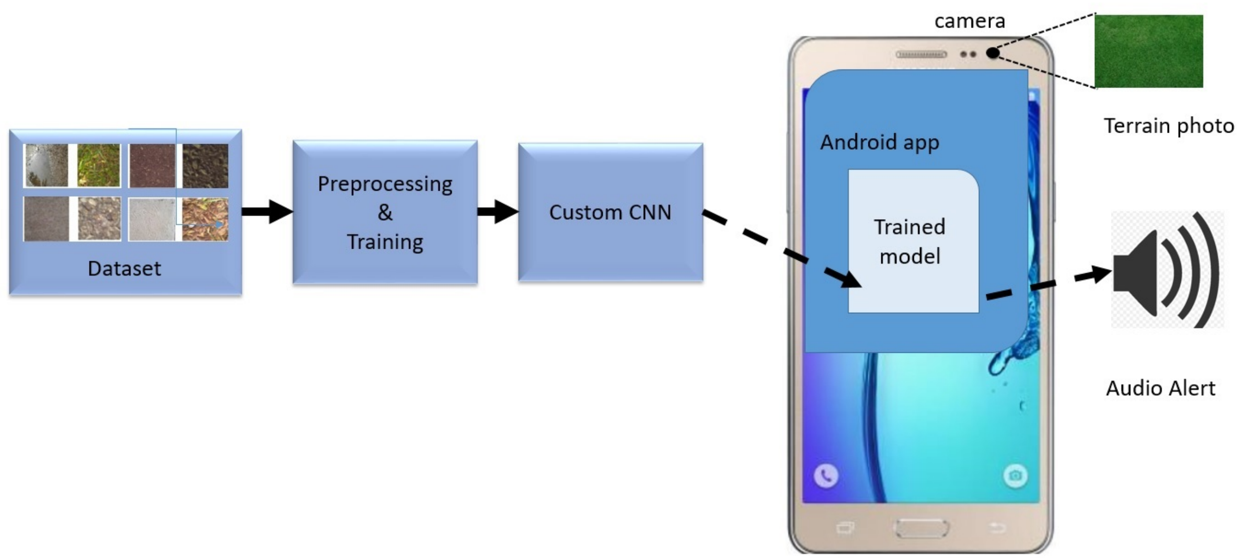

Fig. 1. Flow diagram of the proposed system for ground terrain detection.

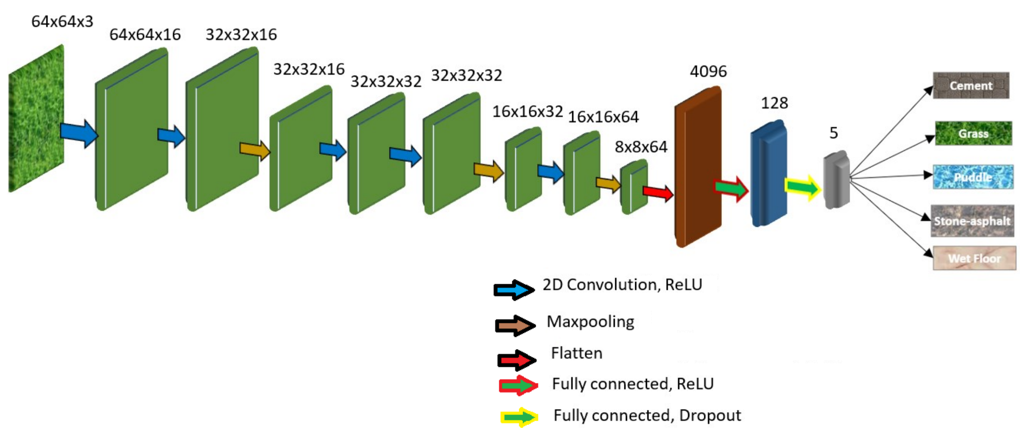

Fig. 2. Architecture of the proposed custom CNN for identifying ground terrain.

TABLE I. LAYER STRUCTURE DETAILS OF MODIFIED CUSTOM CNN

Layer	Input	Kernels	Output	Parameters
Conv1	64 × 64 × 3	16	64 × 64 × 16	448
Conv2	64 × 64 × 16	16	64 × 64 × 16	2320
Maxpool1	64 × 64 × 16	-	32 × 32 × 16	-
Conv3	32 × 32 × 16	32	32 × 32 × 32	4640
Conv4	32 × 32 × 32	32	32 × 32 × 32	9248
Maxpool2	32 × 32 × 32	-	16 × 16 × 32	-
Conv5	16 × 16 × 32	64	16 × 16 × 64	18496
Maxpool3	16 × 16 × 64	-	8 × 8 × 64	-
Flatten	8 × 8 × 64	-	4096	-
Dense1	4096	128	128	524416
Dense2	128	-	5	645

of outdoor ground surfaces, such as concrete, grass, and asphalt-covered stone. Additionally, it incorporates imagery depicting wet patches and puddles. Our dataset draw images from the GTOS-Mobile dataset [18] for ground terrain information, while images featuring puddles and damp floors were sourced from the IEEE Stagnant Water dataset [19]. Furthermore, we enriched the dataset with video footage recorded using a mobile device.

A. GTOS-Mobile Dataset

The GTOS-Mobile dataset [18] encompasses 81 videos that showcase terrain categories similar to those found in the GTOS dataset. These videos were captured using a handheld mobile phone, introducing a variety of lighting conditions and viewpoints. The dataset comprises an extensive collection of 100,000 images distributed across 31 distinct classes. After a meticulous curation process we extracted 6066 frames by employing a temporal sampling rate of approximately 1/10th of a second. However, the emphasis is placed solely on the crucial ground terrain categories. Owing to concerns surrounding image clarity, data pertaining to sand and soil from the GTOS-Mobile dataset is deemed less dependable, prompting its exclusion from consideration. Consequently, for the purpose of outdoor terrain recognition, the analysis narrows down to the inclusion of the cement, grass, and stone-asphalt classes extracted from this dataset. Fig. 3 shows sample images extracted from this data set.

Fig. 3. Images from GTOS-Mobile dataset.

B. Stagnant Water Dataset

The dataset [19] comprises images with dimensions of 256×256 pixels captured using a mobile device. This dataset comprises a total of 1976 annotated photos and encompasses two categories, namely moist surfaces and bodies of water. The dataset is further divided into distinct categories: Indoor, Outdoor, and Raw data. The raw data category comprises unprocessed images, each accompanied by a specified range of resolutions. Puddle detection presents a range of complexities due to fluctuations in lighting conditions, diverse image capture angles, and the presence of reflections on puddle surfaces. To tackle these challenges, approaches such as data augmentation and the collection of images from a diverse range of angles and lighting conditions have been implemented. Fig. 4 shows sample images from the Stagnant Water dataset.

Fig. 4. Images from stagnant water dataset.

Fig. 5. Images from custom dataset.

C. Custom Dataset

A unique dataset was compiled by capturing images using a Redmi Note 7 mobile phone with a 12-megapixel camera. This dataset encompasses 150 images for each distinct category: cement, grass, stone-asphalt, puddle, and wet floor. Images taken by VI persons using mobile phone might not exhibit flawless quality. Intentionally, the custom dataset includes both shaky and low-quality images, mimicking real-world conditions. This deliberate inclusion aims to provide the training process with a more realistic representation of the environment. The images from this custom dataset are depicted in Fig. 5.

The final dataset contains 16763 files belonging to five classes. The images are acquired from the GTOS-Mobile dataset, Stagnant water dataset and images captured using our own mobile device. The training data is split into train data and validation data with a validation split 20%. A total of 12973 files is used as training set, 3243 files for validation, and 759 files as test dataset. Table II shows number of images in each class of custom dataset.

V. Results and Discussion

The CNN model proposed for terrain classification underwent a comprehensive training process to realize the required customized model for accurate terrain detection. The training utilized a specialized dataset encompassing a variety of outdoor terrains, including cement, stone-asphalt, grass, as well as distinct classes for puddles and wet floors. In the

TABLE II. NUMBER OF IMAGES IN EACH CLASS OF CUSTOM DATASET

Class	Train dataset images	Test dataset images
Cement	7550	174
Grass	3100	196
Puddle	1659	114
Stone-asphalt	2370	154
Wetfloor	2084	121

initial stages of training, validation accuracy surpassed 95%, while testing accuracy remained below 70%, indicative of an overfitting scenario. To address this challenge, a combination of techniques including early stopping, dropout, and data augmentation was employed, resulting in a notable reduction of this accuracy gap.

Efforts were directed towards enhancing testing accuracy through layer-wise modifications in the CNN architecture. These refinements led to a remarkable increase in testing accuracy to 98%. The Adam optimizer, with a default learning rate of 0.001, was leveraged to optimize the modified Custom CNN model. Training the model for 300 epochs with Early Stopping set at 50 epochs based on minimum validation loss yielded promising outcomes.

The accuracy and loss curves of the Custom CNN model are visually depicted in Fig. 6, illustrating the progressive improvement achieved during the training process. Meanwhile, Fig. 7 presents the confusion matrix, offering insights into the model's performance across various terrain classes. It is important to note that the observed oscillations in the curves can be attributed to the utilization of an imbalanced dataset in this study. Future work could involve mitigating such oscillations by considering a slightly lower learning rate or implementing exponential decay learning rate strategies. The learning rate, a pivotal parameter in optimization, requires a balanced selection that ensures a trade-off between convergence speed and overshooting tendencies. As evidenced by validation loss values ranging from 0.1 to 0.4, finding this optimal balance remains a crucial consideration.

Further advancements were achieved by enriching the puddle class with additional images sourced from the Stagnant Water dataset. This augmentation elevated the model's validation accuracy to an impressive 98%, while testing accuracy reached 94%, as depicted in Fig. 8. The confusion matrix revealed that, post-enhancement, instances of misclassification emerged wherein puddles were sometimes classified as wet floors. It is worth noting that road puddles and wet floors both fall under the puddle category. The mispredictions of this nature do not significantly impact system functionality, as the system's role is to provide audio feedback advising users to proceed cautiously in case of detected puddles or wet floors.

The proposed CNN model demonstrates the remarkable potential for real-time terrain classification. The results highlight avenues for further refinement, including domain transfer implementations. The integration of this model into practical applications, particularly those necessitating immediate feedback for user safety, underscores its relevance and efficacy in

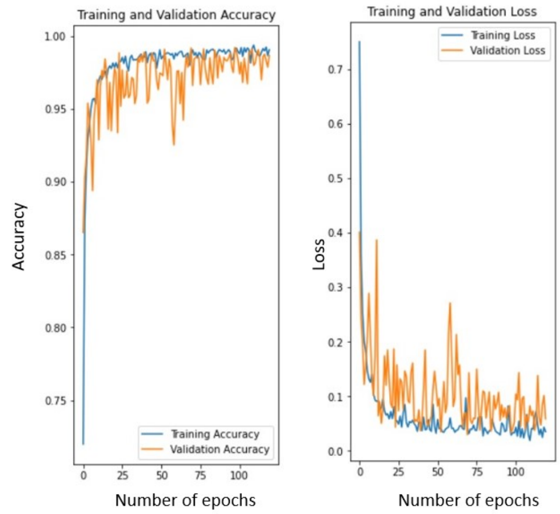

Fig. 6. Accuracy and loss curves of the custom CNN model.

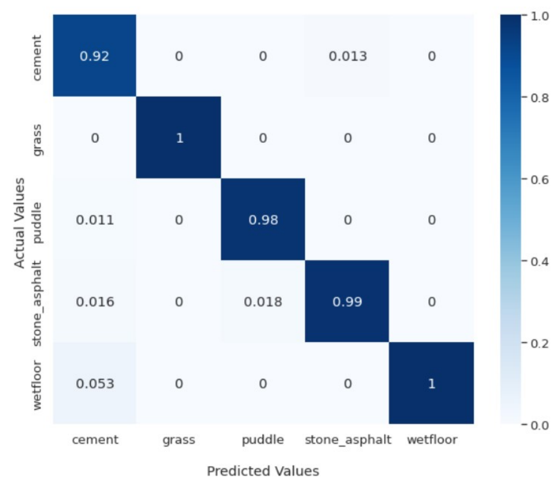

Fig. 7. Confusion matrix of the custom CNN model.

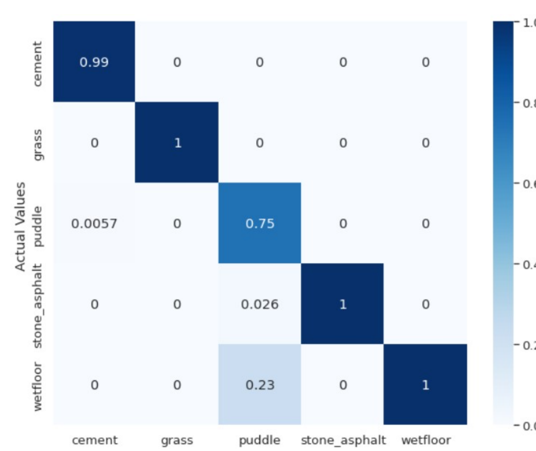

Fig. 8. Confusion matrix of the model after enhancing the puddle dataset.

real-world scenarios.

A. Android Application Results

The central objective behind the proposed system is to enhance the capabilities of conventional robotic white canes by introducing a comprehensive understanding of the surrounding environment. This innovation aims to empower individuals with visual impairments (VIs) by seamlessly integrating global insights into their daily navigation. Rather than augmenting the white cane with additional hardware and algorithmic components, a more user-friendly approach involves encapsulating this functionality within an Android application.

To achieve this goal, a fundamental Android application was developed as a preliminary step to assess the real-time behavior of the proposed deep learning model. This undertaking laid the groundwork for subsequent domain transfer implementations. The transition from a deep learning model stored in the .h5 format to a .tflite model was a pivotal phase. The choice between a quantized or unquantized model depended on size considerations. The deep learning model, which clocked in at 2MB in its original form, maintained this size when converted into an unquantized .tflite model. Opting for a quantized eight-bit .tflite model reduced the size to 500KB. However, this reduction in size through quantization came at a trade-off with accuracy.

The development of the Android application was facilitated by the widely used Android Studio framework, which supports the Java programming language and is well-suited for mobile device applications. After a thorough simulation process, the deep learning model with the lowest validation loss was saved using model checkpointing. To ensure optimal performance on mobile devices, this selected model was converted to an unquantized .tflite model. By doing so, the .tflite model retained a manageable size of 2MB, enabling its seamless integration into the Android application.

The proposed system's core concept revolves around extending the functionality of robotic white canes through a user-friendly Android application. By encapsulating complex insights into a lightweight .tflite model, individuals with visual impairments can gain an enhanced understanding of their surroundings, thereby advancing their autonomy and safety in navigation.

The initial phase of development involved creating a basic version of the Android App, which serves as a foundation for evaluating the system's real-time performance. This App captures images and employs the deep learning model to predict the type of terrain and identify the presence of puddles or wet floors. The primary objective of this basic App version is to gain insights into the system's behavior under real-world conditions.

For testing purposes, the App's user interface includes two buttons, as illustrated in Fig. 9. One button captures an image in real time, while the other allows users to select an image from their gallery. These preliminary App results serve as a foundation for domain transfer evaluations. It is important to emphasize that the basic version of the app is not tailored for individuals with visual impairments (VIs). To address this, an advanced Android app was developed to provide continuous

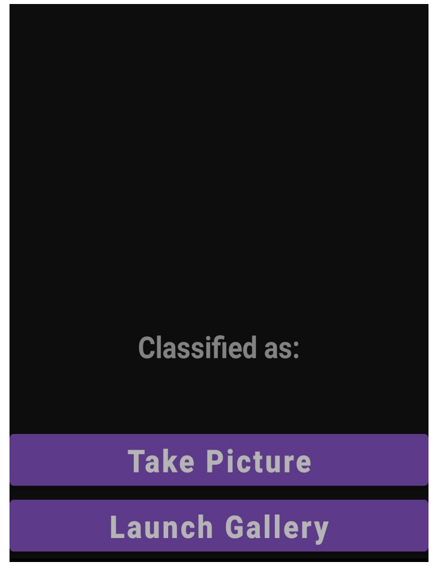

Fig. 9. Android application Interface. Image shows screenshot of the android app, created for testing purposes.

video monitoring and audio feedback, ensuring its suitability and effectiveness for VI users.

VI. Conclusions

In recent decades, advancements in technology have led to the development of robotic and augmented white canes tailored for individuals with visual impairments (VIs). These innovations primarily focus on object detection, obstacle avoidance, and navigation. The major challenge involves providing VI individuals with a comprehensive understanding of their environment to facilitate improved interaction with their surroundings during travel. The central aim of the proposed system is to furnish VI individuals with an encompassing perception of their surroundings. This is achieved through the identification of diverse ground terrains, encompassing cement, grass, and asphalt, along with the detection of road puddles and wet floors. To realize this, a Custom CNN is employed as a deep learning model for a multi-class image classification challenge. The system culminates in conveying vital information to users through audio feedback.

Following an intensive training regimen, the deep learning model achieves commendable performance, boasting a testing accuracy of 94% and an impressive validation accuracy of 98%. To facilitate domain transfer, an Android app was meticulously designed, enabling real-time testing to assess the robustness of the proposed system in various scenarios.

The future trajectory of the proposed system holds significant promise. Plans encompass the expansion of more ground terrain classes to broaden the system's scope and versatility. Additionally, there are intentions to tailor the Android app for individuals with low vision impairments, incorporating continuous video monitoring. This enhancement ensures accurate terrain and puddle identification, even under varying illumination conditions. The App's connection with users will be fortified through intuitive audio feedback, reinforcing its accessibility and usability.

REFERENCES

[1] A. Anwar and S. Aljahdali, "A smart stick for assisting blind people," *IOSR Journal of Computer Engineering*, vol. 19, no. 3, pp. 86–90, 2017.

[2] X. Wang, J. Calderon, N. Khoshavi, and L. G. Jaimes, "Path and floor detection in outdoor environments for fall prevention of the visually impaired population," in *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2022, pp. 1–6.

[3] H. Walle, C. De Runz, B. Serres, and G. Venturini, "A survey on recent advances in ai and vision-based methods for helping and guiding visually impaired people," *Applied Sciences*, vol. 12, no. 5, p. 2308, 2022.

[4] A. Riazi, F. Riazi, R. Yoosfi, and F. Bahmeei, "Outdoor difficulties experienced by a group of visually impaired iranian people," *Journal of current ophthalmology*, vol. 28, no. 2, pp. 85–90, 2016.

[5] A. Wachaja, P. Agarwal, M. Zink, M. R. Adame, K. Möller, and W. Burgard, "Navigating blind people with walking impairments using a smart walker," *Autonomous Robots*, vol. 41, pp. 555–573, 2017.

[6] R. S. Kolhe, K. G. Dhole, P. S. Thakre, P. S. Prasad, P. S. Patel, A. V. Rodge, and F. Akhtar, "Smart stick for the blind and visually impaired people," 2021.

[7] P. Mind, G. Palkar, A. Mahamuni, and S. Sahare, "Smart stick for visually impaired," *Int. J. Eng. Res. Technol*, vol. 10, no. 06, pp. 196–198, 2021.

[8] H. Liu, D. Guo, X. Zhang, W. Zhu, B. Fang, and F. Sun, "Toward image-to-tactile cross-modal perception for visually impaired people," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 2, pp. 521–529, 2020.

[9] C. Xu, K. Wang, K. Yang, R. Cheng, and J. Bai, "Semantic scene understanding on mobile device with illumination invariance for the visually impaired," in *Artificial Intelligence and Machine Learning in Defense Applications*, vol. 11169. SPIE, 2019, pp. 218–226.

[10] F. S. Bashiri, E. LaRose, J. C. Badger, R. M. D'Souza, Z. Yu, and P. Peissig, "Object detection to assist visually impaired people: A deep neural network adventure," in *Advances in Visual Computing: 13th International Symposium, ISVC 2018, Las Vegas, NV, USA, November 19–21, 2018, Proceedings 13*. Springer, 2018, pp. 500–510.

[11] A. A. Elsonbaty, "Smart blind stick design and implementation," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 10, no. 5, 2021.

[12] P. Kozlowski and K. Walas, "Deep neural networks for terrain recognition task," in *2018 Baltic URSI Symposium (URSI)*. IEEE, 2018, pp. 283–286.

[13] S. Zeng, H. Huang, and Z. Shi, "Outdoor terrain recognition based on transfer learning," in *Journal of Physics: Conference Series*, vol. 1846, no. 1. IOP Publishing, 2021, p. 012012.

[14] S. Ghose, P. N. Chowdhury, P. P. Roy, and U. Pal, "Modeling extent-of-texture information for ground terrain recognition," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 4766–4773.

[15] J. Xue, H. Zhang, K. Dana, and K. Nishino, "Differential angular imaging for material recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 764–773.

[16] J. Xue, H. Zhang, and K. Dana, "Deep texture manifold for ground terrain recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 558–567.

[17] K. Yang, L. M. Bergasa, E. Romera, R. Cheng, T. Chen, and K. Wang, "Unifying terrain awareness through real-time semantic segmentation," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, pp. 1033–1038.

[18] J. Xue, H. Zhang, and K. Dana, "Deep texture manifold for ground terrain recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 558–567.

[19] S. Bhutad and K. Patil, "Dataset of stagnant water and wet surface label images for detection," *Data in Brief*, vol. 40, p. 107752, 2022.