Soil friction coefficient estimation using CNN included in an assistive system for walking in urban areas

Oleksiy Gensytskyy, Pratyush Nandi, Martin J.-D. Otis, Clinton Enow Tabi, Johannes C. Ayena

Abstract: We present a smartphone-based solution to help visually impaired people autonomously navigate in urban environments. Contrary to previous works in the literature, the newly proposed system in this paper aims to determine the coefficient of friction (COF) of a soil for aiding in the safe movement of blind and visually impaired people (BVIP). Through convolutional neural networks (CNNs) trained to output this measure, our new investigation then offers the possibility of predicting a risk of falling in their next step by determining the maximum static friction force of the ground. Indeed, a commercial smartphone's camera captures the video of the ground and sends frame as inputs to the CNN model for image segmentation and COF computation. We validated our proposed model in real experiments carried out on 8 types of soils, while also experimenting different CNN models and different optimizers. The proposed ResNet50 CNN-based system provides an accuracy of 96% in the classification of soil type, enabling to guide visually impaired persons. Combined with the associated COF of the soil type, it's possible to estimate a risk of fall (stick or slip) for the next step in the front of the user from the past measured interaction force (using an instrumented insole) between the soil and the sole. Traditional navigation approaches do not consider the soil characteristics such as COF to guide blind and visually impaired person.

Keywords: Smartphone, Risk of fall, Deep learning, Computer vision

1. Introduction

Vision impairment can impact daily activities, such as walking and navigating in a crowded environment due to the inability to clearly understand the surrounding environment (National Academies of Sciences & Medicine, 2017). To overcome this situation, researchers are developing several aiding methods and tools for blind and visually impaired people (BVIP). Among them, the cane and guide dog are the most popular assistants used to navigate both indoors and outdoors. However, they can be expensive and only work under certain conditions. Thus, navigation, particularly in an urban environment, remains a problem that requires more efficient solutions. Indeed, falls, trips, slips and accidents with external objects such as vehicles are common concerns for visually impaired people and can sometimes be hazardous, dangerous and potentially fatal, if not stressful for the walker (Freeman et al., 2007). Therefore, developing new approaches becomes vital in order to motivate the BVIP to continue interacting more easily with their social environment. Thanks to the advancement of the technology, numerous navigation and smart assistive systems (Bal et al., 2020; Fernandes et al., 2019; Islam et al., 2019) have been developed for safe indoor and outdoor walking activities. Computer vision systems and machine learning models are used for purposes, such as obstacle detection (Lin et al., 2014) and pedestrian tracking (Ess et al., 2008), which could help BVIP to better navigate in an urban environment.

Even though neural networks (NN) have made stronger detection capabilities while also understanding the user's surroundings, ignoring the effect of the ground on which the BVIP will walk, would not be practicable for the efficiency of the assistive navigation solutions. For example, Fong et al. (2009) reported that, when the coefficient of friction of the surface drops below 0.41, human starts to walk carefully to adapt to slippery surface. To address these disadvantages in navigation systems, we are therefore estimating the coefficient of friction (COF) of the soil the person is walking on, using a deep learning approach. COF is a dimensionless parameter that is defined as the ratio between force and normal force (Blau, 2001). It depends on the nature of the materials and surface roughness. This coefficient labelled μ, measures the amount of friction existing between two surfaces. In this study, the COF are assessing on 8 types of soils. Indeed, as the BVIP have no way to predict the maximum static friction force of the ground on their next step, this new information could help in creating an assistive system that alerts the user of a possible risk of fall on their next step. In fact, it is known that foot-ground contact is the important part of the forward dynamic biomechanical models.

To achieve this goal, in this paper, we introduce a new approach, which combines a convolutional neural network (CNN), most effective among all current deep learning architectures, with computer vision on a smartphone similar to (Shadi et al., 2019). Following a review of the research and technologies that have been used to assist walking, particularly for BVIP and the modern techniques used for soil classification in an image, we describe the primary contribution of this paper, a neural network-based system used to determine the COF of the soil present in an image in both indoor and outdoor environments using our new database presented by Gensytskyy et al. (2021).

2. Related works

In this section, a brief overview of modern navigation systems for vision impaired people are presented. Then, results of different soil classification algorithms are discussed and compared.

2.1. Assisted navigation for blind and visually impaired people

Various assisted navigation aids, based on the sensory systems, hardware configuration, physical setup, data inference techniques and user interface, are available for BVIP. For walking assistance, navigation technologies can be mainly categorized into visual imagery (IP camera, RGB-D camera, Microsoft Kinect, etc.) and non-visual (BLE beacons, Ultrasonic, IR, etc.) systems, map-based methods, systems with 3D sound and smartphone-based solutions (El-Taher et al., 2021; Kuriakose et al., 2022; Simões et al., 2020). For example, Aladrén et al. (2014) used a consumer-grade RGB-D camera to obtain both visual and depth information for range-based floor segmentation. Their system aimed to provide obstacle-free paths for the visually impaired when walking. They point out that their proposed method has only given good results in indoor environments. Based on RFID technology, some studies (Kahraman & Turhan, 2022; Sáez et al., 2021; Yamashita et al., 2017) demonstrated in real environment that it is possible to navigate safely in complex settings. Two types of data (Bai et al., 2019; Plikynas et al., 2020; Siriboyina & Thadikemalla, 2021) such as inertial sensors and smartphone' camera (Croce et al., 2019) or smartphone and IP camera (Chaccour & Badr, 2016) can also be combined to acquire more accuracy. Even if it is important to highlight the advantages and disadvantages of each different technology available in the literature, this is outside the scope of this research. However, it is known they differ from each other in terms of inaccuracies due to indoor disturbances, availability, comfortability and portability, energy consumption, installation cost and high maintenance (Kuriakose et al., 2022; Van Haute et al., 2016).

In the recent years, machine learning algorithms integrated in such technologies approaches have advanced the field of navigation systems for vision impaired people (Bhandari et al., 2021; Lo Valvo et al., 2021). Specifically, studies employing deep-learning architectures have shown tremendous performance (Schmidhuber, 2015) in mastering the surrounding environment while providing safe navigation compared to the traditional approaches (Bauer et al., 2020; Manjari et al., 2020; Parikh et al., 2018). Elgendy et al.(2021) have used markers and applied deep learning to help visually impaired people walking more easily from one point to the next. Their proposed system was able to find the shortest path to the destination that people with visual impairment should follow. Of the many methods proposed in the literature involving deep learning architectures (Tapu et al., 2020), wearable systems, such as the use of smartphone, seem to be the most promising tools to help BVIP to continue easily, efficiently and in autonomous way their activities in unknown environments (Budrionis et al., 2022; Kuriakose et al., 2020). Usually, in the form of a single device, it has simpler architecture and dynamic requirements. Indeed, in the last decades, smartphones have received much more attention since they have various capabilities including processing power, integrated cameras, and various sensors. Furthermore, they are portable and affordable devices that are already in everyone's pocket. Thus, these possibilities allow researchers to propose more efficient new applications. As suggested in (Alwi & Ahmad, 2013), a compact and easy-to-use system like a smartphone is recently accepted as an interesting choice for an outdoor navigation system for BVIP. After comparing different classifier models, DeepLabV3+ was the model they chose for semantic image segmentation. In the same way, after evaluating many deep learning algorithms, Lin et al. (2017) exploited the Faster R-CNN and YOLO to perform systematic object recognition. Despite the CNNs are the most effective among all current deep learning architectures, their results for obstacle recognition implemented with smartphone application were 60%. Gamal et al. (2020) utilized Google Maps application on a smartphone to plan a route from the current position to the final destination. In their study, CNN networks are employed to guide and help the user to maintain a straight course along the path. Since visually impaired people need assistance for both indoors and outdoors, Saleh et al. (2019) proposed a deep-learning approach implemented on a mobile device to detect obstacles such as pedestrians and cars. Another type of wearable devices such as smart glasses (Ali Hassan & Tang, 2016; Chang et al., 2020), shoe (Raja & Santhosh, 2021), cap exploited IR receivers (Islam et al., 2018), etc., are also designed to fit different parts of the body.

After carefully investigating the literature, we chose a vision-based method with a smartphone camera to establish a navigation system for BVIP, which incorporates a semantic-segmentation process. Indeed, semantic segmentation role's is to find meaningful regions in the processed images and then assign them to a specific class based on the architecture of each pixel (Zhang et al., 2020). However, despite the current extensive research in the literature, the algorithms and technologies related to the use of smartphone and deep learning are still facing issues. The current techniques are highly cost ineffective since the user cannot perceive the type of ground which can represent an important risk of falling. Thereby, in this paper, we are focusing on an assistive system for walking intended primarily for outdoor use in an urban environment. While the studies presented in (Ayena et al., 2015) and (Shadi et al., 2019) aim to investigate obstacle detection and safe-path generation, here we are looking closely for the properties of the ground by estimating its coefficient of friction (COF). Contrary to many works in the literature, distance and location of objects are therefore not the type of information that we are looking forward in this new kind of approach. By exploiting the available techniques in the literature, we are also focusing on a soil classification problem. Since our contribution covers this issue, the next section reviews the current state of soil classification algorithms.

2.2. **Soil classification**

Soil classification is the process of determining the type of soil present on an image. Nowadays, NNs are widely used to classify and locate different types of soils for engineering (Srivastava et al., 2021), geographical (Sefrin et al., 2020) and agricultural (Li et al., 2018) purposes. Inazumu et al*.(2020)* developed a deep learning based model to make decisions on soil classification for three types of soil, namely, sand, clay and gravel, using image recognition with a CNN. Their results showed that generalization performance could be improved by securing enough data for learning. Furthermore, the samples used in their dataset did not differentiate if the soil had a presence of water and were also pictured in a transparent plastic cup, with no external environment or significant noise present other than lightning. Demonstrating that with enough data and image processing, successful classification of a complex image such as soil type is possible in a controlled scene, but remain challenging when the image scene in question is disparate and noisy such as the background outdoor environment present in the image. Hence, we consider the nature of soil, which considers environmental hazards present at the time of computing the COF. A solution to secure enough learning data could be to use image augmentation techniques as the soil is essentially a repeating-colored texture in an image. Transformations like translation and rotation can be used for multiplying image count (May, 2019) and will be looked further upon. Support vector machine (SVM) models are also used for soil classification (Rao et al., 2016; Srunitha & Padmavathi, 2016) and can be combined with a CNN at the output layer instead of using CNN-SoftMax for instance. CNN-SVM models can help to improve accuracy in some classification problems, but it is not always the case like Agarap, A. (2017) observed when comparing both architectures using the Fashion-MNIST dataset (Xiao et al., 2017), as CNN-SVM showed inferior testing accuracy than the CNN-SoftMax model (90.7% and 91.8% respectively).

To our best knowledge, no work has previously designed a CNN to fully automate an assistive technique for the BVIP in order to give to them a perception of the type of soil they are walking on. A novel vision-based method is therefore proposed in this paper to follow design requirements for real-time. It integrates both the smartphone-based technology mentioned above to push back their limitations while exploiting their benefits.

3. Suggested methodology and process

In this section, we describe the methodology of our work for creating a better assistive system. Mainly, the whole process of the COF extraction from an image is explained thoroughly, which consists of the image preprocessing and augmentation, and the training on the CNN model.

3.1. Wearable device and assistive system

For the proposed system, a camera, a fast-enough processing unit which is packaged in both small, portative, and battery-powered is required. As reported above, a smartphone is a good choice for a vision assistive system as it has a camera, battery, and computing unit. Modern smartphones such as an iPhone 12 have enough computing power for the purpose of this research (Apple). The back camera of the smartphone is intended to be facing the walker's front point of view (POV) for the system to be able to see the ground, which is the soil the person is walking on. The smartphone is not required to be held in hands when the person is walking. For ease of use and comfortableness, the smartphone could be placed vertically inside a shirt pocket, with the back camera still facing the front walker's point of view.

The smartphone is therefore the wearable device used in this paper, and it is essentially used as a computer vision system. For each walking step, it will automatically take a picture, process, and analyse the taken image in order to estimate the COF of the soil present in the picture and then evaluate a risk of fall based on a simple mechanical model. To compute the risk of fall, the interacting force between sole and soil, including the angle of the foot, are an input to estimate if the sole stick or slip on the soil. This study uses an instrumented insole using Bluetooth (BLE) to transmit these data (interacting force and foot angle) to the smartphone.

A mobile application can be programmed for the smartphone which will automatically repeat this process while the person is walking. The proposed assistive system can be seen in Figure 1 where as proposed in (Patarot et al., 2014; Rodrıguez & Corchado, 2020; Stimac et al., 2015), a smart belt equipped with sensors such as a camera could also act as the computer vision system and replace the need of a smartphone.

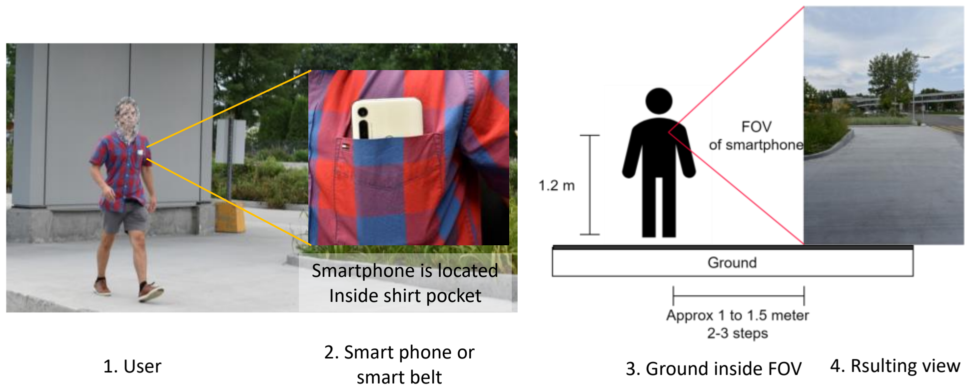


Figure 1 : Suggested acquisition system (Note: FOV: field of view) to predict next walking step risk of fall

Table 1 : Static coefficient of friction between rubber and different soils

Soil type	Dry COF	Wet COF	References
Ice	0.1	n/a	(Lo Valvo et al., 2021; Physicsteacher.in, 2019)
Snow	0.3	n/a	(hypertextbook.com, 2007)
Gravel	0.35	n/a	(Noon, 1994)
Grass	0.36	n/a	(Engineersedge.com, 2023)
Epoxy flooring	0.55	n/a	(El-Sherbiny et al., 2012)
Concrete	0.6	0.45	(Engineersedge.com, 2023; TheEngineeringToolbox.com)
Wood	0.7	n/a	(Herbert-Wertheim-College-of-Engineering, 2023)
Asphalt	0.9	0.75	(Engineersedge.com, 2023)

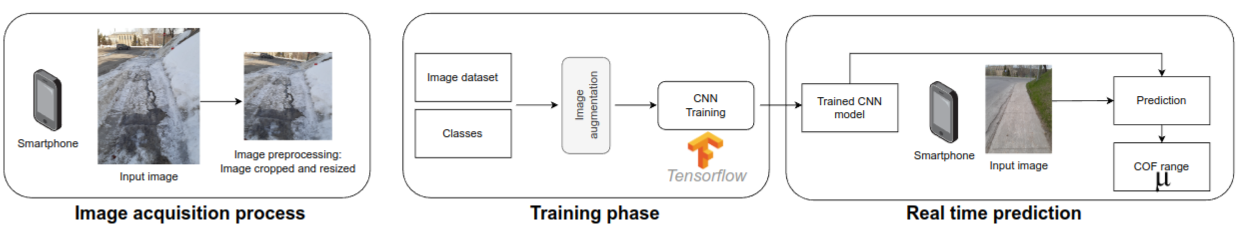


Figure 2 : Data piping for image acquisition, training, testing, and real-time acquisition using a smartphone

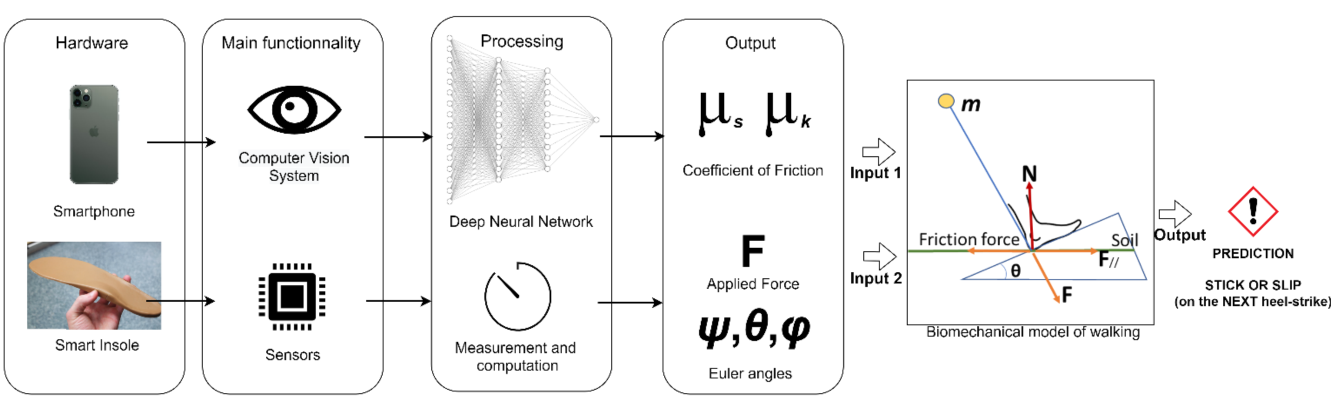


Figure 3: Illustration of proposed process to evaluate a risk of fall in real-time using a smartphone and an instrumented insole

3.2. Methodology and process

What we intend to do in this paper is to determine the COF of the ground (or soil) present on an image taken from a smartphone camera. The smartphone is placed vertically inside a shirt pocket, with the back camera facing the front POV of the walker. The COF is a value between 0 and 1 which describes the ratio of the force of friction between two bodies and the force pressing them together. Table 1 contains different static COFs for common types of soils with rubber on dry and wet conditions. These values will be used for the determination of the COF of any type of soil present in an image.

Our proposed system aims to classify the soil type present in an image using a Deep Neural Network and then lookup its COF from the Table 1 (Figure 2). The processing of the image is done with a CNN that is used as an image classifier and the smartphone as the camera. A trained model will be available on the smartphone and then used by an internal application to detect the type of soil that is present in the picture. Figure 2 presents the block diagram of our process which is described in the next subsections. In this approach, the pictures of different types of soils are fed into our soil classification deep neural network model. The COF of the soil is eventually deduced from the output of this classification and eventually becomes input to the walking biomechanical model enabling a risk of fall estimation. Of course, to compute and evaluate a risk of fall, the system needs the position of the foot on the ground. An inertial measurement unit and force sensors are integrated in an instrumented insole to measure orientation of the foot according to the ground and applied force as shown in Figure 3. By using the COF computed using an image classifier, it is possible to compute a risk of fall based on friction cone: is the foot slipping or sticking on the ground? This approach gives two phases in the gait where the risk should be higher: (1) *the first contact of the heel on the ground (heel strike)* and (2) *the propulsion just before the toes leave the ground* (*toe off*). This risk of fall level is stated as a potential risk of fall. Of course, a fall is not only based on the stick/slip of the foot, but also on the balance recovery capacity. Higher risk could occur in the propulsion phase or heel strike phase in the gait cycle. The mechanical model (biomechanical model of walking shown in Figure 3) needs an history of the previous measured forces at the heel strike and propulsion to estimate a risk based on the COF in the front of the user (next heel strike).

3.2.1 Datasets

In the beginning of this research, we explored the option of using available datasets such as COCO, Cityscapes and ETH pedestrian as shown in Table 2. However, we figured that we could not use any of them because nor the context nor the aimed use of the dataset was in line with the end purpose of our proposed assistive system. First, none of the presented datasets had annotated soil type on their images. Second, not all images have a clear ground (or soil) area present in them. For example, in the COCO dataset, some images only have an object present such as a motorcycle or a building, with no sufficient area of ground on the image for an algorithm to segment and detect its soil type if there is any.

As an alternative of going through a large available dataset and filtering the images that are of use for our system and lose tremendous time on annotating the type of soil on each of one of them, we decided to create our own dataset which is publicly accessible (Gensytskyy, 2021). Our dataset contains 493 images and 8 different types of soils (Table 3). While our new dataset is not large as the COCO dataset, the context and the POV of our images better represent the end-usage of our proposed assistive system and have a clear view of the ground.

Accordingly, the proposed CNN model is trained with our own image dataset. Each image was taken outside in good lightning conditions on different locations and different types of soil (Table 3) in the region of Saguenay, Quebec, Canada and were taken from the POV of the height of a shirt pocket, as one of the intended uses of the assistive system. Figure 4 shows some example images from our dataset along with Table 3 which shows the number of images per class.

Table 2. Other open datasets

Dataset	Usage	Problem
COCO (Common Objects in Context) (Lin et al., 2014)	Large-scale object detection and segmentation	Not all images have ground or soil present.
Cityscapes (Cordts et al., 2016)	Large-scale dataset for semantic image segmentation	Images are taken in street scenes from the point of view the hood of a vehicle. The only type of soil present is pavement.
ETH pedestrian (Ess et al., 2007)	Pedestrian detection and tracking	The only type of soil present is pavement.

Table 3. The different types of soil and the corresponding number of images collected.

Classes	**Number of images**
Asphalt	89
Concrete	80
Gravel	58
Ice	40
Snow	68
Wood	34
Grass	90
Epoxy flooring	34
Total	**493**

Figure 4: Images of soils with different coefficients of friction from our dataset

The types of soils that are taken into consideration in our system are: snow, ice, gravel, concrete, asphalt, grass, wood and epoxy flooring (Figure 4). Therefore, we are aiming to classify the soil type the person is walking on, and then output its according COF from table 1.

We are aiming to detect the soil type using a CNN, and then output its corresponding COF to a biomechanical model of walking for friction force calculation. The next subsections of the paper present the detailed implementation of the suggested algorithm including iterative improvements.

3.2.2 Segmentation and feature selection

An image can contain various visual components, and all together these components complexify the overall information that needs to be analyzed. Thus, segmentation techniques are often applied on images to extract the relevant data. There are two main ways to segment the ground from a vertical image as the use proposed in our system: 1) create a bounding box in the image and 2) use semantic segmentation.

For the first way, it could be to create a bounded-box area (x, y, width, height) on the bottom of the image. This area would contain a part of the ground in the image. The downside of using this method is that by creating a bounded box, if the ground is not in the designed area, further processing would fail. This problem is alleviated as the system could correct itself at the next frame input. This method is simple and works because the POV of the camera faces forward and the height position makes so that the ground will always be visible from its POV and that the ground will always be at the bottom of the image. The second way to look at the problem is to use semantic segmentation techniques. CNN-based segmentation can be used to isolate the ground and various objects in an image for analysis at a later stage (Cordts et al., 2016; Li et al., 2018). In this case, the picture is firstly taken and is then segmented with a CNN. Each pixel is labeled as a class (e.g., ground, car, pedestrian). The position of the ground-labeled pixels is used as reference to create a dynamic-bounded box inside. This method would reduce error in further processing as the input data would be more accurate and well segmented. While semantic image segmentation aims to segment single images, semantic video segmentation aims to use the temporal continuity of multiple images taken to improve classification results (Gadde et al., 2017; Shelhamer et al., 2016).

The proposed way in (1) is simple, and while a bounded box area could lead to errors, it is not that likely as the ground is always at the bottom of the image and the system could correct itself at the next input. The idea in (2) is more complex as it would require a 2-stage CNN for instance, but it could improve classification results as segmentation would be more accurate. However, we know that a small error on the estimated COF is not enough to change the risk of fall.

3.2.3 Image Pre-processing

In this paper, we rely on the images that were collected manually while walking in different locations on different types of soil. We apply three pre-processing steps as illustrated in Figure 6. In the following order, we have:

1) *Cropping*: In order to consider only the ground on which the person is walking on, a cropped section of the image was considered. For cropping, a square of 20% of the overall image area was considered in order to avoid noise (car, stair, cloud, sky, etc.) coming from other objects. The idea was to take lower half of the image and crop center square of the image. After considering the cropping, the accuracy decreased as compared to the images that were taken initially before cropping. One hypothesis for why this happens is that the CNN model considered environment as well while calculating the COF, like trees or nearby houses and cars

2) *Rotating and Resizing*: This technique is used to first rotate a random angle and resize the image into a specified size. This size was chosen at 224x224 pixels format as shown in Figure 5 and mainly depend of the computation capacity available (memory allocation capacity/overflow, processing power, RAM/VRAM size).

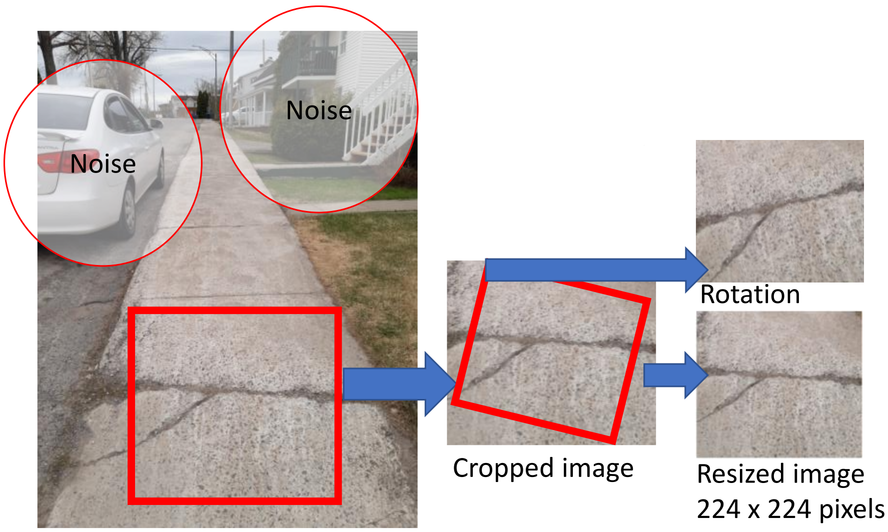


Figure 5 : Image pre-processing: cropping, rotating, resizing

3) *Normalizing*: We used standard deviation and mean which were computed using ImageNet pretrained model.

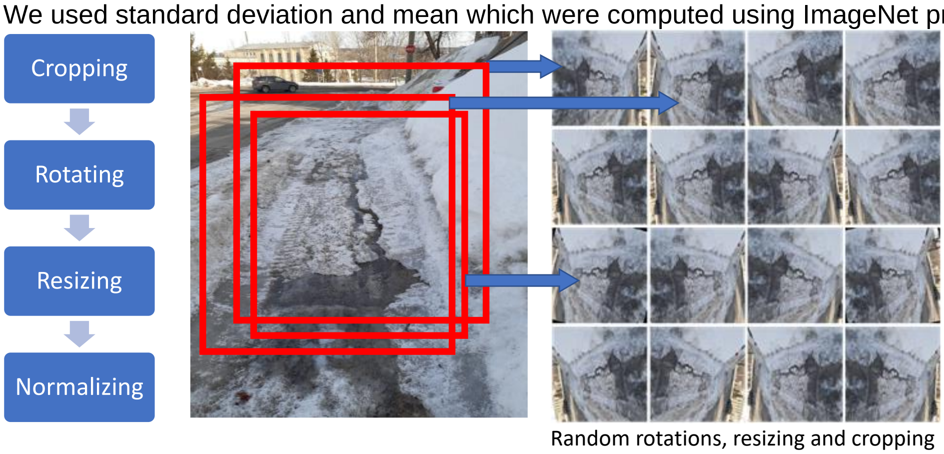


Figure 6: Pre-processing steps applied on the collected images.

3.3. Neural Network model

3.3.1 CNN

In a neural network, we have an input layer, hidden layer, and output layer. If there are multiple hidden layers, then this network is considered as a deep neural network. The hidden layer captures the non-linear relationship between the input and output layer. A Convolutional Neural Network (CNN) is a special type of neural network which works very well with data that is spatially connected. Spatial information refers to information having location-based relation with other information. A group of pixels is used for feature prediction instead of using a single pixel value. For our use case, the spatial and translation invariance of the CNN will be useful to classify different classes of types of soils. Residual Learning is a technique used in neural networks that reduces the error rate on neural networks with increased layer depth, and reduces the complexity of the network. For example, ResNet50 is a deep residual convolutional neural network (CNN) made of 50 layers.

Furthermore, we have investigated three iterations to identify the COF as described below: 1) using straight images including all picture taken by the smartphone (no cropping, no rotating, no resizing), with small CNN, 2) using image augmentation with deeper CNN and 3) using image augmentation with ResNet50**.** ResNet is the winner of classification challenge in the ILSVRC-2015 competition, and it is characterized by a very deep network with 50 / 101 / 152 layers. It was developed by He et al. (2016). This architecture was formed to defeat challenges in deep leaning training because it takes quite a lot of time and is limited to a certain number of layers. The advantage in ResNet models is that, compared to other architectural models, its performance does not decrease even though the architecture is getting deeper. Besides, computation calculations are made lighter, and the ability to train networks is better. Thus, in our study and for the next process, we kept the model 3. The suggested architecture of the ResNet50 used in this study is presented in Figure 7.

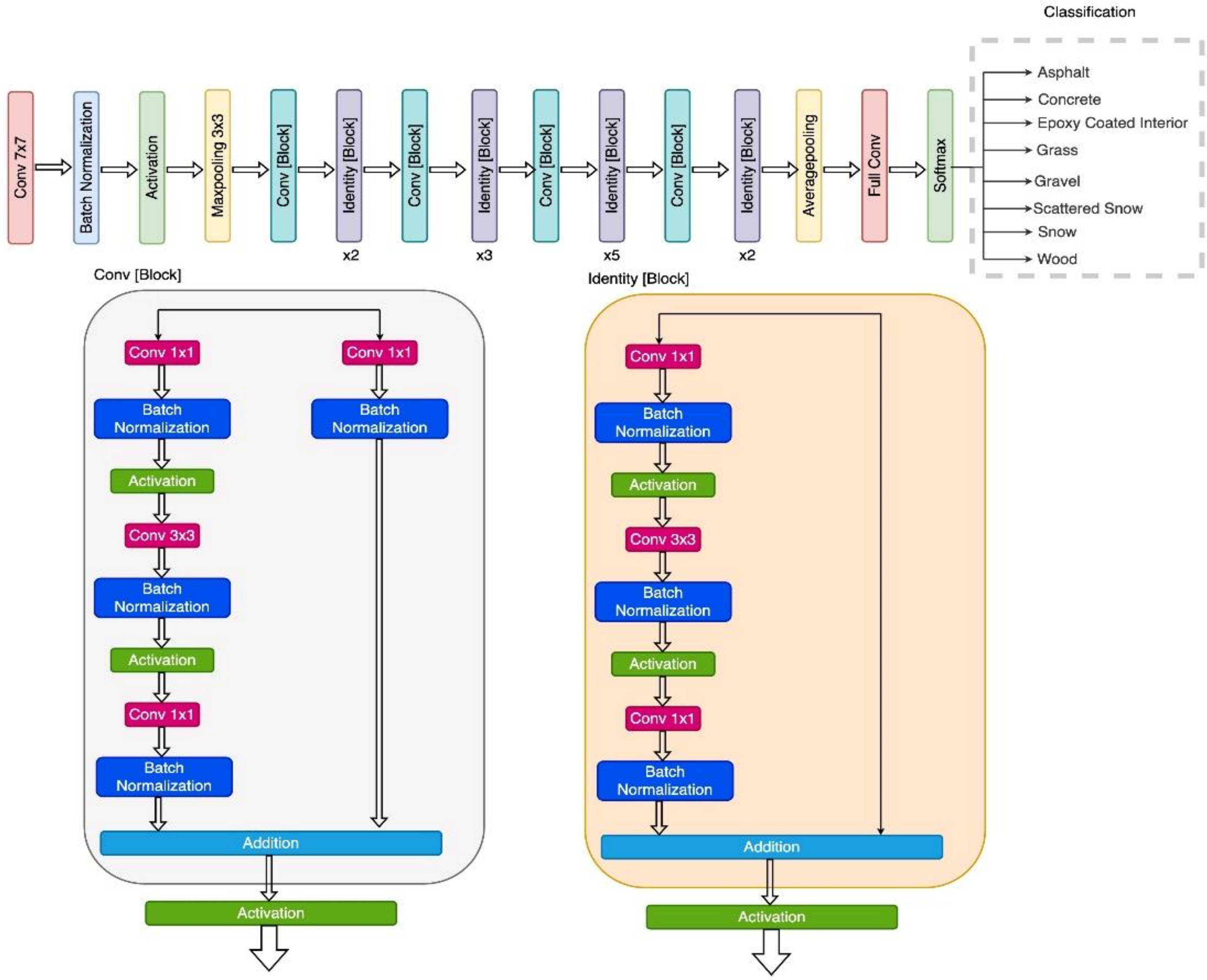


Figure 7: Suggested ResNet50 architecture.

An additional custom layer was added at the output of the ResNet50 model to fit the output classes dimensions. A dropout layer was also added to prevent the overfitting of the training data. On top of this, we made our custom network which consisted of (Figure 8):

- Linear layer
- Relu activation function
- Dropout layer
- Linear layer
- LogSoftmax activation function

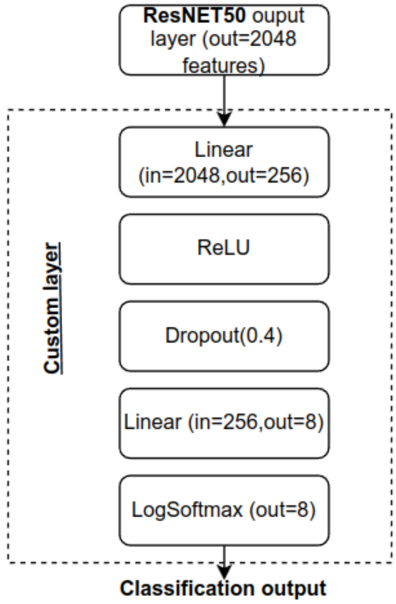


Figure 8: Custom output layers added at the end of the ResNet50 model.

3.3.2 Image Augmentation

Image augmentation is a processing technique used to increase the amount of image data in a given dataset. It consists of altering an image in multiple ways by applying random transformations in an image such as cropping, zooming, rotation or noise, among others, to create multiple altered copies of the original image. This can considerably multiply the number of available data for the training process in the CNN and thus improve model results when given a small dataset size (Shorten & Khoshgoftaar, 2019). Nevertheless, image augmentation is not always successful for every type of dataset for machine learning, depending on the nature of data, the transformations being made and the end classification goal in the CNN. Since our dataset is small and neural networks require a big amount of data, we applied the image augmentation technique in the training phase (Figure 5-6) of the CNN models to improve results. New images were created on the fly to feed in our neural network while in training step. The augmentation transformations below were applied in training phase to increase our data set:

- Random resizing and cropping
- Random rotation by 15 degrees
- Adding color jitters
- Random horizontal flip
- Center crop
- Normalizing pixels

Figure 9 presents an example of an augmented Gravel image from our dataset. We used transfer learning for this project. Transfer learning is the process of repurposing knowledge from one task to another. Essentially if we look at it from a modelling perspective it means using a model trained on one data set and fine-tuning it for use with another.

3.3.3 Experimental evaluation

As described above, image augmentation technique was used since the proposed model is trained on a small dataset. The data is sent in batches and is trained for 100 epochs. We have also used a condition wherein if the model is trained with parameters/weights that has less training accuracy than the previous epoch then the training is stopped there. The parameters/weights which were acquired in the previous epoch are considered. Table 4 represents the hyperparameters used in the proposed model to optimize the performance of classifier (iteration 3). The commonly used optimization algorithm is the gradient descent algorithm which can optimize the deep learning model. We use an adaptive moment estimation (ADAM) optimizer which calculates adaptive learning rates for each weight. The loss function used is negative log likelihood loss.

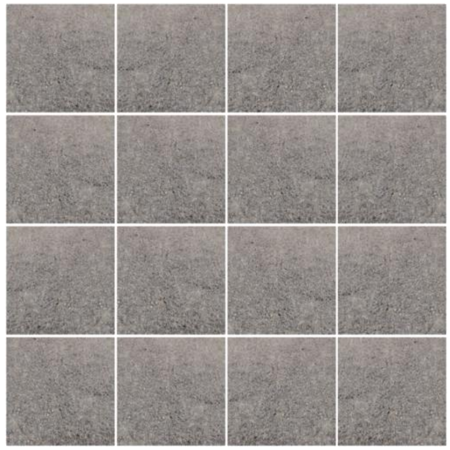


Figure 9 : Example of image augmentation of gravel image

Table 4. Hyperparameters of CNN model.

Hyperparameters	**Value**
No. Of Epochs	100
Batch Size	64
Optimizer	ADAM
Loss Function	Negative Log Likelihood Loss

Table 5. Confusion matrix's results of the different types of soils used.

		Target (actual) classes							
		Asphalt	**Concrete**	**Epoxy Coated Interior**	**Grass**	**Gravel**	**Scattered Snow**	**Snow**	**Wood**
Predicted (output) classes	Asphalt	19	0	0	0	1	0	0	0
	Concrete	1	15	0	0	0	0	0	0
	Epoxy Coated Interior	0	0	7	0	0	0	0	0
	Grass	0	0	0	18	0	0	0	0
	Gravel	1	0	0	0	10	0	0	0
	Scattered Snow	0	0	0	0	1	7	0	0
	Snow	0	0	0	0	0	0	14	0
	Wood	0	0	0	0	0	0	0	7

Table 6. Recognition results metrics.

Class	**Precision**	**Recall**	**F1-score**
Asphalt	0.95	0.95	0.95
Concrete	1	1	1
Gravel	0.91	0.91	0.91
Scattered snow	1	1	1
Snow	1	1	1
Wood	1	1	1

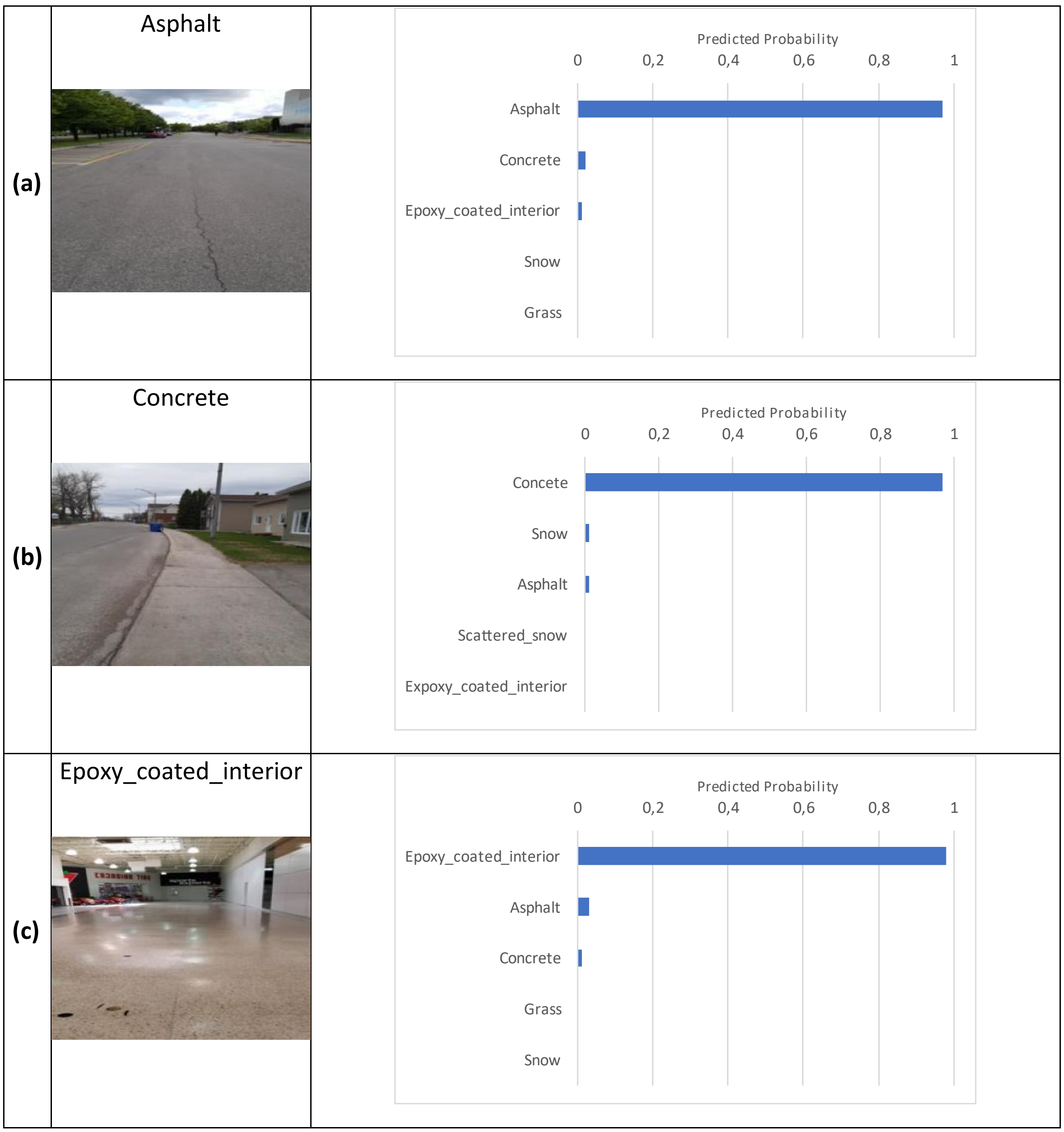
4. *Results*

4.1. Image classification

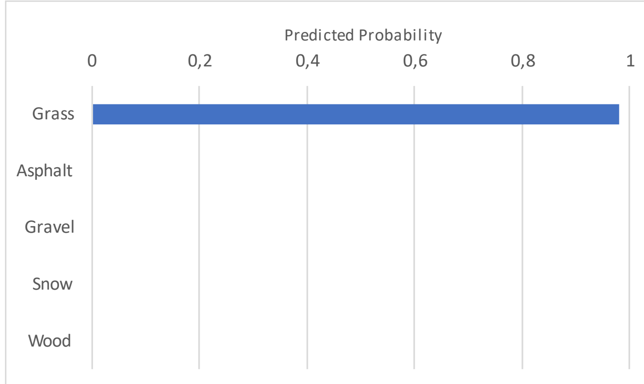
In the following, we report the results related to the confusion matrix as well as the classification probability from our proposed CNN model which generated a good classification rate.

A confusion matrix (a two-dimensional matrix with one dimension indexed with actual class and another dimensioned indexed with the class assigned by the classifier) summarizes the classification performance of the classifier. It was used on the test data for analyzing the classifier's performance. Table 5 represents the confusion matrix of the prediction outcome of the test data group on the trained model. The horizontal rows represent the actual class, while the vertical columns represent the number of predictions corresponding to each possible soil in the set. The image dataset was split into 60% for training purposes and 20% each for testing (20%) and validation (20%) groups. Table 6 presents the metrics for the performance of the suggested classifier.

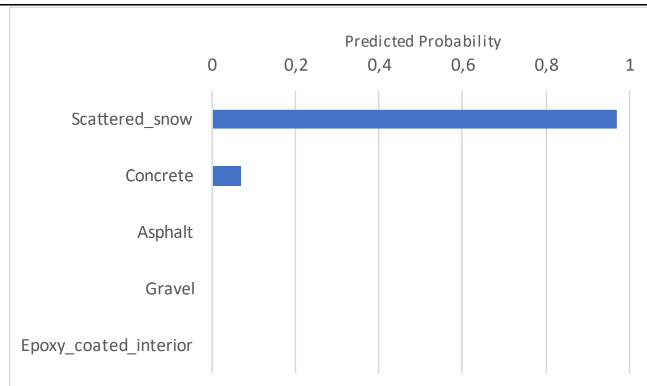
We also measured the probability with which the other classes are being predicted. In this way, we remark that there is a co-relation between different classes although they might look different as shown in Figure 10. As an example, Figure 10(h) depicts the low probability with which Asphalt class is getting detected though it's a Gravel class.



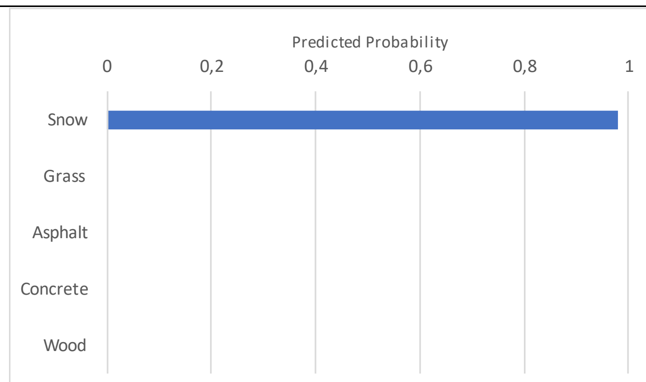
(d) Grass



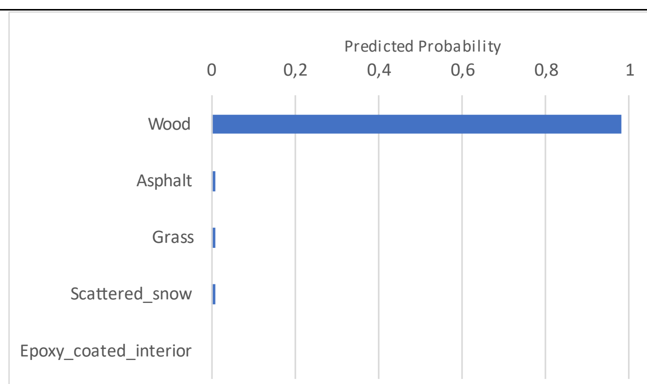
(e) Scattered_snow



(f) Snow



(g) Wood



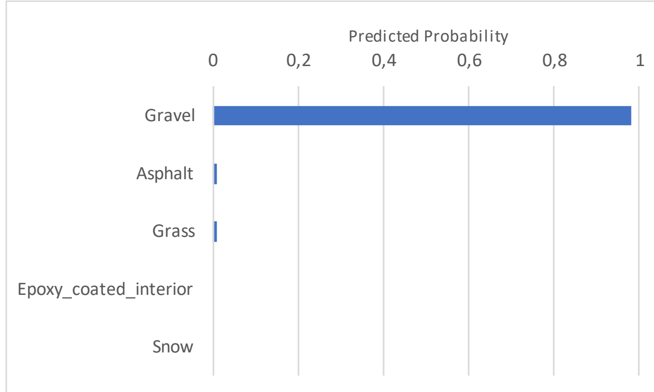
(h)	Gravel 	

Figure 10 : Prediction results on test images of various soil types

4.2. Risk of fall evaluation

The biomechanical model is designed primarily to estimate the risk of fall R during walking for the next heel strike according to the soil type the user is walking on and the past forces applied on the ground at the heel strike and proposition phase (toe off) in the gait cycle. In Figure 11, the free body diagram is presented as the model to compute a risk of fall. In this figure, the risk becomes high as the COF decreases and as the angle theta (θ) increases. First, one hypothesis considers that right standing posture (while walking or not) do not represent any risk of fall as shown in Figure 11(a). Then, the second hypothesis considers heel strike, shown in Figure 11(b), and toe off as the two mains gait phases with higher risk of falling. The suggested model considers these two phases.

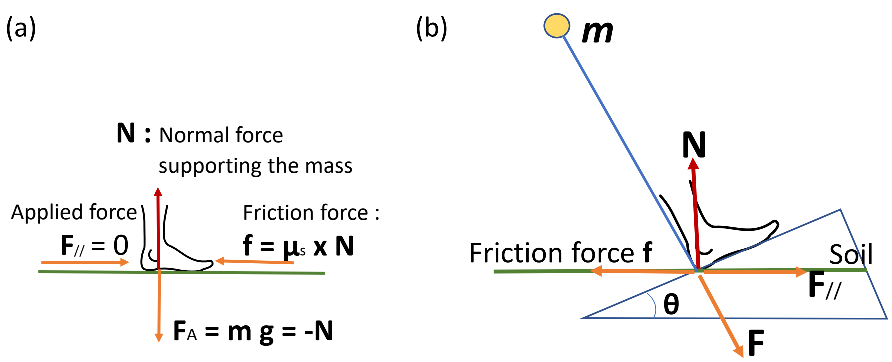


Figure 11: Free body diagram for computing risk of fall: (a) without risk (b) heel strike model with higher risk of falling

The risk of falling is computed as the fraction between the sum of the forces applied at the point of contact between the heel-ground and the frictional force, as described in equation (1) and (2):

$$\text{Risk Of Fall} = \frac{\sum \text{Applied forces(at the point of contact betweeen the heel and the ground)}}{\text{Friction force}} \tag{1}$$

$$\text{Risk of fall} = \frac{mgsin\theta cos\theta}{\mu_s N} = \frac{mgsin\theta cos\theta}{\mu_s mgcos\theta} = \frac{sin\theta}{\mu_s} \tag{2}$$

where *m* and *g* represent the mass of the person and gravity acceleration respectively, μ_s is the static coefficient of friction, while N is the applied normal force on the ground as shown in Figure 11.

On a flat surface, when the foot is parallel to the ground, the angle θ is 0 rad. On a heel strike to the ground, the angle is typically no higher than 0.7 rad (in the Figure 11) as it would be an abnormal walking position or may indicate that the user has slipped or felt. For the heel strike, the angle is positive. For the support phase, the angle is 0 and the angle becomes negative for propulsion phase (until toe off).

A Risk of Fall (ROF) lower than 1 is a good indication that the foot adheres to the ground, hence, the person maintains his balance. As the ROF becomes closer to 1, the risk of fall is significant, as the ratio of applied force to static friction force becomes 1, the shoe is at the limit between static and kinetic friction, thus making the user more prone to slipping and falling. The applied force is measured by the insole in real time using force-sensitive resistors (FSR). It is then possible to submit an alert (vibrotactile or sound alert) to the person before a potential fall. Figure 12 shows the relation between the COF and theta variables when computing the COF.

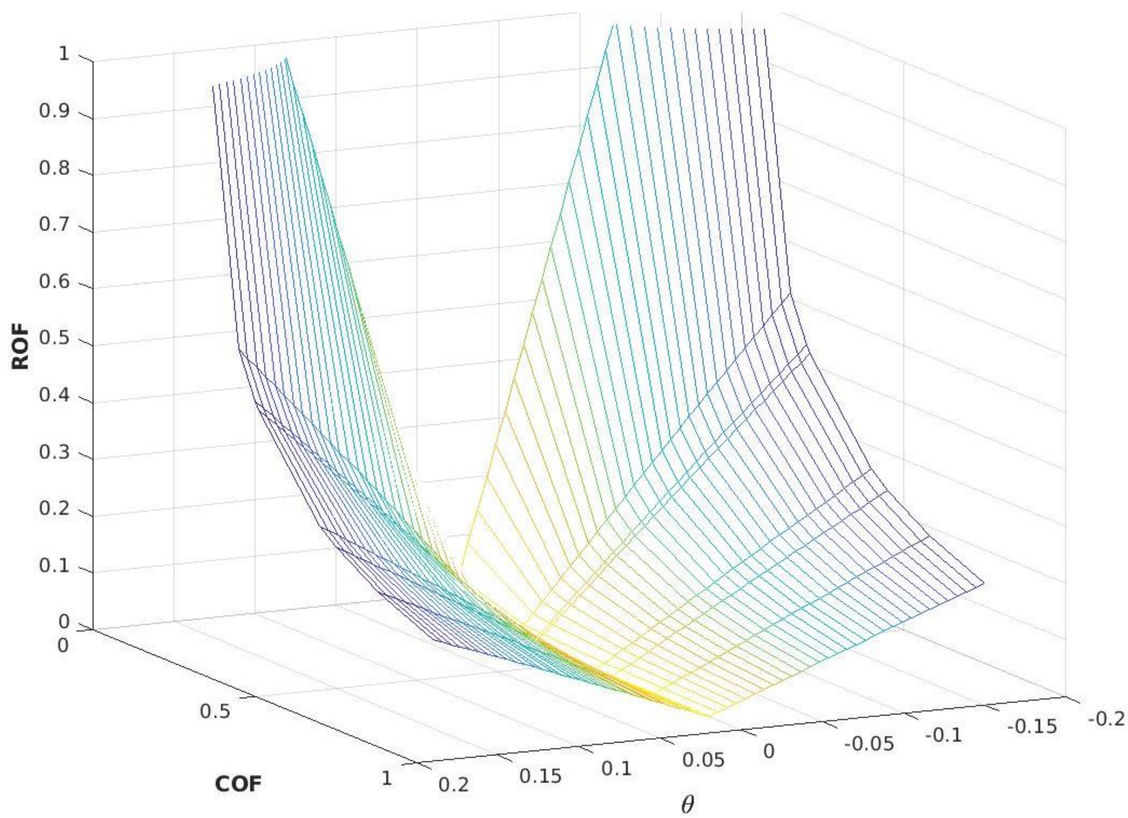


Figure 12: Risk of fall by foot angle (in radians) and COF

5. Discussion

Previously developed systems for the BVIP assume that the user already knows the path to follow to their final destination (Elgendy et al., 2021; Gamal et al., 2020; Shadi et al., 2019). The research is more focused on the design of a system for indicating the presence of obstacles in the front, left and right directions around the surroundings. However, when a BVIP is visiting an environment, such as walking for the first time on a new type of soil, this may not be the case and therefore special assistance may be required. Thereby, our present study particularly emphasized the usefulness of the automatic determination of the COF of the type of soil the BVIP is walking on. The outcomes of our evaluation show that determining the COF is found to be significantly important to enhance the development of new solutions in navigation system area. Indeed, the COF of the soil can provide additional information to guide BVIP.

Once our proposed CNN model had been trained, we evaluated the results considering confusion matrix represented in Table 5, as the third iteration of improvement. From the results, it is evident that this model used with ResNet50, despite its more complex architecture, provided the best results. In most cases, the type of soil was correctly identified in more than 90% of images. Another important metric we evaluated is the probability with which the other classes are being predicted (Figure 10). This further demonstrates the good performance of the model 3. We can conclude that the model 3 using ResNet50 with a custom output layer is the best solution for the identification of the COF in an image.

In this paper, it is proposed that the user is wearing one smart insole, not two, since the walking gait is highly similar for both feet. However, it’s the same process using another insole: the insole is used to measure the force and the COF is predicted with the same smartphone. This is due to simplicity and economic constraints. For instance, each smart insole needs to have its battery charged, be configured and connected to a smartphone via Bluetooth or BLE. Of course, it's possible to use one insole in each foot (the equation is the same), but it is reasonable to use only one instrumented insole and a second dummy (with no electronics).

6. Limit of the study

The type of soil classification technique used in this paper is based on CNN that automatically detects features in an image by training on an image dataset. The image dataset is separated in classes (labels), and the algorithm learns the features (in the training phase) that makes an image similar or different from the other classes. It is widely known that neural networks trained with a bigger dataset will yield better results or in some cases overfit (saturate) the network. In this research project, publicly available datasets that have big dataset quantity are not appropriate for our project. Some datasets are cited with an explanation why they are not appropriate. That being said, this paper has developed and suggested a new dataset that is publicly available. If the data quantity of our dataset was raised, the algorithm would perform better; it is a question to take the pictures, to make sure they are acceptable to be used as training data and to pre-process them. However, in order to address the image data quantity problem, image augmentation techniques were used to make the training dataset bigger.
The other problem that arises with the pictures in the dataset is that they can contain unwanted elements or noise that can have a negative effect on the accuracy of the CNN. The image can contain part of a car, a bin or many types of soil in the same vicinity which makes the detection of the soil type complex. So further analysis to automatically remove unwanted elements of the input image would help the accuracy of the detection algorithm. In this project, a simple approach is presented to crop the input images to only keep the bottom-half of the picture in such to remove unwanted elements of the image, since only the part of the image that is important to the algorithm is the ground, which contains the features that are detected by the convolutional neural network.
Finally, the last issue is related to the detection of some important physical properties such as a layer of water over the ice. It is well known that the layer of water as an impact on the COF. However, the current suggested system is not able to detect such situation.

7. Conclusions

In this paper, we exploited ResNet50 to configure a device for the visually impaired when walking in an urban environment. The main contribution of the paper is to use the soil characteristics to detect a potential slipping of the sole on the soil type. The navigation system is realized using a commercial smartphone (iPhone 12). The results observed show that the image taken from the smartphone's camera can be associated to a COF corresponding to a type of soil. Then, a risk of fall could be estimated according to the estimation of the COF in the front of the user for the next walking steps. This investigation aids in the safe execution of outdoor daily activities of visually impaired people on a specific type of soil. The accuracy of the CNN output can be decreased due to many factors in the user environment (intensity of light -night, day, etc.-, physical properties of the ground, etc.) which will be taken as future work. Also, for better performance, the proposed model should be trained for larger data sets containing more different types of soils.

As future work, the estimation of some ground properties needs to be studied in more detail. As an example, ice is very difficult to detect. However, ice itself could stick with the sole. The issue is more related to the fine layer of water on the ice as suggested in wet COF in Table 1. Therefore, other sensors could be used to detect this situation such as temperature and humidity sensors. It is also possible to use the data coming from meteorological data. A fusion and prediction algorithms could be used to validate the possibility to estimate a fine layer of water over the ice. Then, it could be possible to adjust the COF according to the prediction. As the issue is to predict the risk of fall for the next few walking steps, using the sensors located in the insole is not considered as they only provide information in the current step, not for the next steps.

Next, to improve the COF estimation, we aim to increase our dataset size by adding more pictures for the CNN training process. Currently, our dataset counts about ≈ 500 images, and some classes have less images than others, mainly due to time constraint in the project and environmental changes in the physical location of the team (winter, summer). Other than increasing its size, we will add more pictures of different types of soils in new locations, lightning and angles to improve its estimation of the COF in different, never-seen before contexts by the CNN. Also, to reduce the case of unwanted noise in the pictures, we hope to use modern image segmentation techniques such as semantic segmentation to increase the quantity of data to be used for the detection of the soil type, as our current technique is simple but limits the number of pixels to be used as input for soil classification process in the CNN.

We are also investigating the extension of the proposed system. In particular, we can consider the use of other wearable devices, e.g., smart-watch devices, in order to generate a more comfortable experience for the BVIP. In order to better assess the usability of our proposed system, we plan to recruit visually impaired human subjects to conduct experiments in different types of soils. This developed system will encourage the BVIP to become more active by increasing their mobility and their quality of life.

8. References

Agarap, A. F. (2017). An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification. *arXiv preprint arXiv:1712.03541*.

Aladren, A., López-Nicolás, G., Puig, L., & Guerrero, J. J. (2014). Navigation assistance for the visually impaired using RGB-D sensor with range expansion. *IEEE Systems Journal*, *10*(3), 922-932.

Ali Hassan, E., & Tang, T. B. (2016). Smart glasses for the visually impaired people. Computers Helping People with Special Needs: 15th International Conference, ICCHP 2016, Linz, Austria, July 13-15, 2016, Proceedings, Part II 15,

Alwi, S. R. A. W., & Ahmad, M. N. (2013). Survey on outdoor navigation system needs for blind people. 2013 IEEE Student Conference on Research and Developement,

Apple.). *Iphone 12 specifications*. https://www.apple.com/ca/fr/iphone-12/specs/

Ayena, J. C., Zaibi, H., Otis, M. J.-D., & Ménélas, B.-A. J. (2015). Home-based risk of falling assessment test using a closed-loop balance model. *IEEE transactions on neural systems and rehabilitation engineering*, *24*(12), 1351-1362.

Bai, J., Liu, Z., Lin, Y., Li, Y., Lian, S., & Liu, D. (2019). Wearable travel aid for environment perception and navigation of visually impaired people. *Electronics*, *8*(6), 697.

Bal, D., Tusher, M. M. I., Rahman, M., & Saymon, M. S. R. (2020). NAVIX: a wearable navigation system for visually impaired persons. 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI),

Bauer, Z., Dominguez, A., Cruz, E., Gomez-Donoso, F., Orts-Escolano, S., & Cazorla, M. (2020). Enhancing perception for the visually impaired with deep learning techniques and low-cost wearable sensors. *Pattern recognition letters*, *137*, 27-36.

Bhandari, A., Prasad, P., Alsadoon, A., & Maag, A. (2021). Object detection and recognition: using deep learning to assist the visually impaired. *Disability and Rehabilitation: Assistive Technology*, *16*(3), 280-288.

Blau, P. J. (2001). The significance and use of the friction coefficient. *Tribology International*, *34*(9), 585-591.

Budrionis, A., Plikynas, D., Daniušis, P., & Indrulionis, A. (2022). Smartphone-based computer vision travelling aids for blind and visually impaired individuals: A systematic review. *Assistive Technology*, *34*(2), 178-194.

Chaccour, K., & Badr, G. (2016). Computer vision guidance system for indoor navigation of visually impaired people. 2016 IEEE 8th international conference on intelligent systems (IS),

Chang, W.-J., Chen, L.-B., Hsu, C.-H., Chen, J.-H., Yang, T.-C., & Lin, C.-P. (2020). MedGlasses: A wearable smart-glasses-based drug pill recognition system using deep learning for visually impaired chronic patients. *IEEE Access*, *8*, 17013-17024.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. Proceedings of the IEEE conference on computer vision and pattern recognition,

Croce, D., Giarre, L., Pascucci, F., Tinnirello, I., Galioto, G. E., Garlisi, D., & Valvo, A. L. (2019). An indoor and outdoor navigation system for visually impaired people. *IEEE Access*, *7*, 170406-170418.

El-Sherbiny, Y., Hasouna, A., & Ali, W. (2012). Friction coefficient of rubber sliding against flooring materials. *ARPN J. Eng. Appl. Sci*, *7*(1), 121-126.

El-Taher, F. E.-Z., Taha, A., Courtney, J., & Mckeever, S. (2021). A systematic review of urban navigation systems for visually impaired people. *Sensors*, *21*(9), 3103.

Elgendy, M., Sik-Lanyi, C., & Kelemen, A. (2021). A novel marker detection system for people with visual impairment using the improved tiny-yolov3 model. *Computer Methods and Programs in Biomedicine*, *205*, 106112.

Engineersedge.com. (2023). *Coefficient of Friction Equation and Table Chart*. Retrieved 2021-08-05 from https://www.engineersedge.com/coeffients_of_friction.htm

Ess, A., Leibe, B., Schindler, K., & Van Gool, L. (2008). A mobile vision system for robust multi-person tracking. 2008 IEEE Conference on Computer Vision and Pattern Recognition,

Ess, A., Leibe, B., & Van Gool, L. (2007). Depth and appearance for mobile scene analysis. 2007 IEEE 11th international conference on computer vision, Fernandes, H., Costa, P., Filipe, V., Paredes, H., & Barroso, J. (2019). A review of assistive spatial orientation and navigation technologies for the visually impaired. *Universal Access in the Information Society*, *18*, 155-168.

Fong, D. T.-P., Hong, Y., & Li, J.-X. (2009). Human walks carefully when the ground dynamic coefficient of friction drops below 0.41. *Safety science*, *47*(10), 1429-1433.

Freeman, E. E., Munoz, B., Rubin, G., & West, S. K. (2007). Visual field loss increases the risk of falls in older adults: the Salisbury eye evaluation. *Investigative ophthalmology & visual science*, *48*(10), 4445-4450.

Gadde, R., Jampani, V., & Gehler, P. V. (2017). Semantic video cnns through representation warping. Proceedings of the IEEE International Conference on Computer Vision,

Gamal, O., Thakkar, S., & Roth, H. (2020). Towards intelligent assistive system for visually impaired people: Outdoor navigation system. 2020 24th International Conference on System Theory, Control and Computing (ICSTCC),

Gensytskyy, O. B. A., Yessine; Otis, Martin J.-D. (2021). *Image dataset of various soil types in an urban city*. https://doi.org/https://doi.org/10.5683/SP3/R8FDJB

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition,

Herbert-Wertheim-College-of-Engineering. (2023). *Friction coefficients between different wheel/tire materials and concrete*. Retrieved 2021-08-05 from https://mae.ufl.edu/designlab/Class%20Projects/Background%20Information/Friction%20coefficients.htm

hypertextbook.com. (2007). *Friction coefficients*. Retrieved 2021-08-05 from https://hypertextbook.com/facts/2007/TabraizRasul.shtml

Inazumi, S., Intui, S., Jotisankasa, A., Chaiprakaikeow, S., & Kojima, K. (2020). Artificial intelligence system for supporting soil classification. *Results in Engineering*, *8*, 100188.

Islam, M. I., Raj, M. M. H., Nath, S., Rahman, M. F., Hossen, S., & Imam, M. H. (2018). An indoor navigation system for visually impaired people using a path finding algorithm and a wearable cap. 2018 3rd International Conference for Convergence in Technology (I2CT),

Islam, M. M., Sadi, M. S., Zamli, K. Z., & Ahmed, M. M. (2019). Developing walking assistants for visually impaired people: A review. *IEEE Sensors Journal*, *19*(8), 2814-2828.

Kahraman, M., & Turhan, C. (2022). An intelligent indoor guidance and navigation system for the visually impaired. *Assistive Technology*, *34*(4), 478-486.

Kuriakose, B., Shrestha, R., & Sandnes, F. E. (2020). Smartphone navigation support for blind and visually impaired people-a comprehensive analysis of potentials and opportunities. International conference on human-computer interaction,

Kuriakose, B., Shrestha, R., & Sandnes, F. E. (2022). Tools and technologies for blind and visually impaired navigation support: a review. *IETE Technical Review*, *39*(1), 3-18.

Li, W., Dong, R., Fu, H., & Yu, L. (2018). Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks. *Remote Sensing*, *11*(1), 11.

Lin, B.-S., Lee, C.-C., & Chiang, P.-Y. (2017). Simple smartphone-based guiding system for visually impaired people. *Sensors*, *17*(6), 1371.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13,

Lo Valvo, A., Croce, D., Garlisi, D., Giuliano, F., Giarré, L., & Tinnirello, I. (2021). A navigation and augmented reality system for visually impaired people. *Sensors*, *21*(9), 3061.

Manjari, K., Verma, M., & Singal, G. (2020). A survey on assistive technology for visually impaired. *Internet of Things*, *11*, 100188.

May, P. (2019). Improved image augmentation for convolutional neural networks by copyout and CopyPairing. *arXiv preprint arXiv:1909.00390*.

National Academies of Sciences, E., & Medicine. (2017). *Making eye health a population health imperative: Vision for tomorrow*. National Academies Press.

Noon, R. K. (1994). *Engineering Analysis of Vehicular Accidents*. CRC Press LLC.

Parikh, N., Shah, I., & Vahora, S. (2018). Android smartphone based visual object recognition for visually impaired using deep learning. 2018 International Conference on Communication and Signal Processing (ICCSP),

Patarot, A., Boukallel, M., & Lamy-Perbal, S. (2014). A case study on sensors and techniques for pedestrian inertial navigation. 2014 International Symposium on Inertial Sensors and Systems (ISISS),

Physicsteacher.in. (2019). *Friction coefficient*. Retrieved 2023-07-11 from https://physicsteacher.in/2019/10/13/friction-coefficient/

Plikynas, D., Žvironas, A., Budrionis, A., & Gudauskis, M. (2020). Indoor navigation systems for visually impaired persons: Mapping the features of existing technologies to user needs. *Sensors*, *20*(3), 636.

Raja, L., & Santhosh, R. (2021). Experimental study on shoe based navigation system for the visually impaired. *Materials Today: Proceedings*, *45*, 1713-1716.

Rao, A., Gowda, A., & Beham, R. (2016). Machine learning in soil classification and crop detection. *International Journal for Scientific Research and Development*, *4*(1).

Rodrıguez, S., & Corchado, J. M. (2020). Smart Belt Design by Naıve Bayes Classifier for Standard Industrial Protection Equipment Integration. *Int. J. Artif. Intell*, *18*, 186-201.

Sáez, Y., Montes, H., Garcia, A., Muñoz, J., Collado, E., & Mendoza, R. (2021). Indoor navigation technologies based on RFID systems to assist visually impaired people: A review and a proposal. *IEEE Latin America Transactions*, *19*(8), 1286-1298.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.

Sefrin, O., Riese, F. M., & Keller, S. (2020). Deep learning for land cover change detection. *Remote Sensing*, *13*(1), 78.

Shadi, S., Hadi, S., Nazari, M. A., & Hardt, W. (2019). Outdoor navigation for visually impaired based on deep learning. Proc. CEUR Workshop Proc,

Shelhamer, E., Rakelly, K., Hoffman, J., & Darrell, T. (2016). Clockwork convnets for video semantic segmentation. Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14,

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, *6*(1), 1-48.

Simões, W. C., Machado, G. S., Sales, A. M., de Lucena, M. M., Jazdi, N., & de Lucena Jr, V. F. (2020). A review of technologies and techniques for indoor navigation systems for the visually impaired. *Sensors*, *20*(14), 3935.

Siriboyina, L. P., & Thadikemalla, V. S. G. (2021). A Hybrid System to Assist Visually Impaired People. *SN Computer Science*, *2*(4), 333.

Srivastava, P., Shukla, A., & Bansal, A. (2021). A comprehensive review on soil classification using deep learning and computer vision techniques. *Multimedia Tools and Applications*, *80*, 14887-14914.

Srunitha, K., & Padmavathi, S. (2016). Performance of SVM classifier for image based soil classification. 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES),

Stimac, P. J., Demar, R. W., Hewitt, G. F., McKenna, M. J., Jordan, E. M., Fordham, M., & Haas III, J. W. (2015). Pedestrian dead reckoning using a novel sensor module that interfaces with modern smart devices. Geospatial Informatics, Fusion, and Motion Video Analytics V,

Tapu, R., Mocanu, B., & Zaharia, T. (2020). Wearable assistive devices for visually impaired: A state of the art survey. *Pattern recognition letters*, *137*, 37-52.

TheEngineeringToolbox.com. *Friction - Friction Coefficients and Calculator*. Retrieved 2021-08-05 from https://www.engineeringtoolbox.com/friction-coefficients-d_778.html

Van Haute, T., De Poorter, E., Crombez, P., Lemic, F., Handziski, V., Wirström, N., Wolisz, A., Voigt, T., & Moerman, I. (2016). Performance analysis of multiple Indoor Positioning Systems in a healthcare environment. *International journal of health geographics*, *15*(1), 1-15.

Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*.

Yamashita, A., Sato, K., Sato, S., & Matsubayashi, K. (2017). Pedestrian navigation system for visually impaired people using HoloLens and RFID. 2017 Conference on Technologies and Applications of Artificial Intelligence (TAAI), Zhang, S., Ma, Z., Zhang, G., Lei, T., Zhang, R., & Cui, Y. (2020). Semantic image segmentation with deep convolutional neural networks and quick shift. *Symmetry*, *12*(3), 427.