2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Indoor/Outdoor Deep Learning Based Image Classification for Object Recognition Applications

Omar Abdullatif Jassim[1]*, Mohammed Jawad Abed[1], Zenah Hadi Saied[2]

[1]Department of Medical Instrumentation Techniques Engineering, College of Al Hikma, University, Baghdad, Iraq.
[2]Department of Medical Laboratory Technologies, Institute of Medical Technology-Al-Mansour, Middle Technical University, Baghdad, Iraq.
*Corresponding Author.

Received 29/11/2023, Revised 02/04/2023, Accepted 04/04/2023, Published 05/12/2023

Abstract

With the rapid development of smart devices, people's lives have become easier, especially for visually disabled or special-needs people. The new achievements in the fields of machine learning and deep learning let people identify and recognise the surrounding environment. In this study, the efficiency and high performance of deep learning architecture are used to build an image classification system in both indoor and outdoor environments. The proposed methodology starts with collecting two datasets (indoor and outdoor) from different separate datasets. In the second step, the collected dataset is split into training, validation, and test sets. The pre-trained GoogleNet and MobileNet-V2 models are trained using the indoor and outdoor sets, resulting in four trained models. The test sets are used to evaluate the trained models using many evaluation metrics (accuracy, TPR, FNR, PPR, FDR). Results of Google Net model indicate the high performance of the designed models with 99.34% and 99.76% accuracies for indoor and outdoor datasets, respectively. For Mobile Net models, the result accuracies are 99.27% and 99.68% for indoor and outdoor sets, respectively. The proposed methodology is compared with similar ones in the field of object recognition and image classification, and the comparative study proves the transcendence of the propsed system.

Keywords: Deep learning, GoogleNet, Image classification, Indoor/outdoor, Transfer learning.

Introduction

The current big data shared across different social media platforms requires powerful tools to be analysed and explored. The rapid development in the field of computer science has created new techniques that are capable of processing large amounts of data, making decisions about them, defining relationships between them, recognising hidden information, etc.[1].

Machine learning (ML), as a computer science field, has been widely used in many data science applications[2], including image classification, clustering, data mining, pattern recognition, etc. However, ML has its limitations, especially when dealing with large datasets.

Deep learning (DL) is one of the most efficient and powerful artificial intelligence (AI) fields that can deal with large data size (image, text, video, etc.)[3].

Image recognition is one of the most common tasks of DL. In image recognition, the image is introduced to the DL model as an input. Then, the feature-pyramid of the input image is built layer by layer using convolution filters. The final feature vector is then classified into the appropriate class[4].

DL is considered a combination of neural networks that have a deeper architecture than the traditional neural networks[5,6]. Due to this deep architecture, DL has high performance and is more capable of dealing with large data size.

Image classification and recognition are the most common task of DL networks. The main idea of this type of DL application is the use of image

2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

datasets. Each dataset is split into training, validation and test sets. The training set is used to train and learn the DL model, while the validation set is used to validate the model during the training process. The test set, on the other hand, is used after the model is trained in order to evaluate its accuracy and performance[7].

ML and DL have many essential applications in biometrics, data mining, image classification, image segmentation, network security, special-needs systems, etc.

According to[8], more than 39 million people around the world are facing a real problem of not recognising their environment nor interacting with other people.

Related Work

For object detection and recognition, many ML and DL systems were designed. Smart systems that have been recently designed to make people do their daily activities easily and effectively[9].

Tapu et al.[10] used the YOLO CNN deep architecture to detect objects in real time in order to help special-needs people in the outdoor environment. This network detects objects like cars, people, pedestrians, etc. They added new classes to the network to be detected, like smartphones and laptops. Their proposed system recognized 30 different objects with 90% accuracy and more than 90% robustness using a dataset of 60 videos. The research focused only on the outdoor environment.

Vijaybahadur et al.[11] designed a virtual assistant to help blind people recognise their surrounding environment. Their proposed methodology used built-in voice recognition and text-to-speech Python libraries to recognised blind voice orders and translate reactions into voices, allowing them to integrate with their environment. The main issue with their study was that there was no validation step. They used built-in systems without any modification, taking into account the special needs of the blind people.

Ephzibah[12] introduced an ML system to assist blind people in their daily life. His proposed system captures the user environment and recognises the components inside the captured image. The proposed methodology was based on the built-in object detection models of the Python "Tensorflow" library. Their study defined if there were closed object without defining its nature; besides, their study had no evaluation process.

Qureshi et al.[13] used the well-known convolutional neural networks (CNN) to design an assistant application. The designed application takes a photo using a mobile then the captured photo is introduced to the CNN, which analyses it and detects the appropriate category (car, person, animal, etc.). The proposed system was evaluated, and the test accuracy was 78%. Their study used a dataset of only 25 individuals (7 of them were completely blind) which is very small dataset.

Al Mamun et al.[14] introduced a new AI system for helping blind people using DL techniques. Their proposed system uses the MTCNN to process and classify the captured images from a mobile camera. The study used the Google API to translate voice into textual messages and the messages back to audible voice. The AI system could receive messages from the blind, recognise its environment, and transfer recognised objects into audible voices. The system was evaluated by 300 participants, but no accuracy or other performance was mentioned.

Chaurasia et al.[15] designed an automated guiding system for blind people based on machine learning techniques. They took into account the indoor environment only (like a school, house, library, coffee shop, mall, etc.). They utilised many computer vision capabilities, like destination detection, voice recognition, and navigation, in order to guide blind people. Their proposed system can be used as a guide for special-needs people to guide them through walking outside. They used the image processing and pixel-based manipulation processes, spending too much time for training and evaluation.

Mone et al.[16] introduced an ML-based computer vision system for visually impaired people. They used the MobileNet deep learning architecture trained on the Coco image dataset in order to detect specific objects and help the user recognise his environment. The main drawback of this study is that they didn't take into account the actual blind needs, but rather used a previously trained deep network on a general dataset without any specification. They should retrain the model on a specific blindness-related dataset. No evaluation process or experimental discussion were mentioned in their research. The size of the used dataset was also unknown.

In a study conducted by Wang et al.[17], reinforcement learning and object detection were employed. They targeted the visually impaired, helping them recognise their environment in an effective way. Target recognition, along with

2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

SceneNet, were used in their research in order to detect objects. They called their model by ObjNet and SceneNet. The experiments were applied to a small dataset of 7 categories, each of which contained about 1000 instances, and the obtained accuracies were 91.3%, 92.1%, and 95.4% for ObjNet, SceneNet and mixed model, respectively. The main limitation of their research was using a small dataset.

Wearable devices for blind people have been proposed in[18]. The designed system had four eye sensors and one wrist sensor. The main part of the device was an Arduino board. The measured distances using the sensors were entered into a decision support based fuzzy logic system in order to recognise the distance of the target object. No image processing was involved in this system. The system was evaluated using participants, but no accuracy or error rate were mentioned.

Visually impaired and blind people were targeted in the study[19]. They proposed a guiding system for those types of people to help them navigate inside stores and markets by recognising the directions, shop entries, exits, menus of goods, and shopping history. The designed system required an internet connection and a smart card. The experiments were conducted with participants. They confirmed that their mobility had been enhanced using this system. No validation accuracy or other error rates were mentioned by this study.

Another similar system was proposed by Dhou et al.[20]. They used ML techniques to help blind and visually impaired people walk and navigate outdoors. The proposed system detected obstacles and helped visually impaired people by using a camera, a smart phone, special sensors, and a digital motion processor in order to analyse movements, objects, and other obstacles on the way. In their study, Nave Bayes, Decision Tree, SVM, and k-Nearest Neighbour ML algorithms were used in their study. The experiments were applied to two different categories (down stairs and hollow pits), and the obtained accuracies were between 99% and 100%.

Many problems were detected in those previous studies. Some previous studies took the indoor environment into account, while others dealt only with the outdoor one. In some studies, the ML and DL models were trained on general and not specific blind-related datasets[16] which will be treated in this study. Some studies didn't evaluate their models [12,14], while others used only a few number of classes. Some studies got low accuracies [10,13]. No one of previous studies discussed the error rates and accuracies of each individual category of indoor and outdoor classes.

In this study, both the indoor and outdoor environments will be taken into account. The main motivation of this study is to take into account all previous studies limitations and try to solve them. The total number of categories is 36, including indoor and outdoor classes, which is bigger than all previous studies. The deep learning models GoogleNet and MobileNet-V2 will be used and trained using these 36 categories and 19905 images (big dataset rather than small datasets (like in some studies)), resulting in powerful DL models. The final model will be tested and evaluated to determine its performance. The discussion will take into account the individual error rate in order to define the best and worst recognized categories of all indoor and outdoor classes.

Materials and Methods

In the current research, the proposed system will take into account the most essential objects and places that people may be concerned about. Moreover, the system will be built in two different environments (indoor and outdoor) so that people can use it wherever they go. Fig 1 describes the proposed system architecture. In the first step, the dataset is collected using different resources in order to take all special-needs individuals' requirements into account. In the second step, the dataset is split into training, validation and test sets. While in the third step, the deep learning GoogleNet and MobileNet-V2 models are trained and validated using the training and validation sets, respectively. The last step includes the evaluation of the proposed methodology using the test set and many performance metrics.

2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

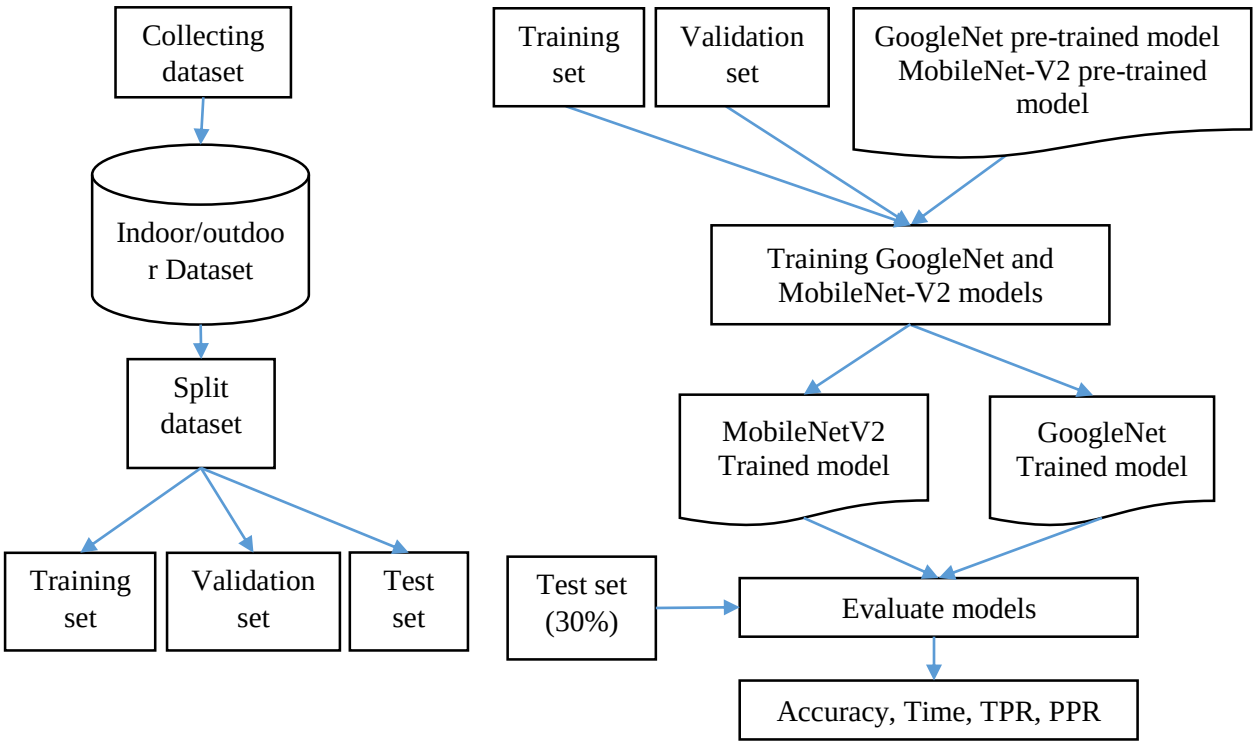


Figure 1. System proposed methodology.

Dataset

Many open source datasets are available for the purpose of object detection. The main concept of this study is to select the most appropriate datasets. Special-needs people deal with specific objects inside or outside their home like food, other people, cars, animals, furniture, etc. Consequently, the categories illustrated in Fig. 2 are suggested.

The Dataset has two main categories (indoor and outdoor). Since the "person" category can be indoor or outdoor, the third category will be the "indoor/outdoor" category.

Indoor category includes 4 main classes and 27 sub-classes (total subclasses) including: Apple, Apricot, Avocado, Banana, Cherry, Cucumber, Kiwi, Lemon, Onion, Orange, Pear, Pepper, Pineapples, Potato, Strawberry, Tomato, Watermelon, Sofa, Table, Chair, Bed, Glass, Mobile, Laptop, Fan and Remote. The outdoor category consists of three main classes, including six sub-classes, which are: cat, dog, flower, motorbike, car and airplane. The mixed indoor/outdoor category has two classes (person and dangerous person).

2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

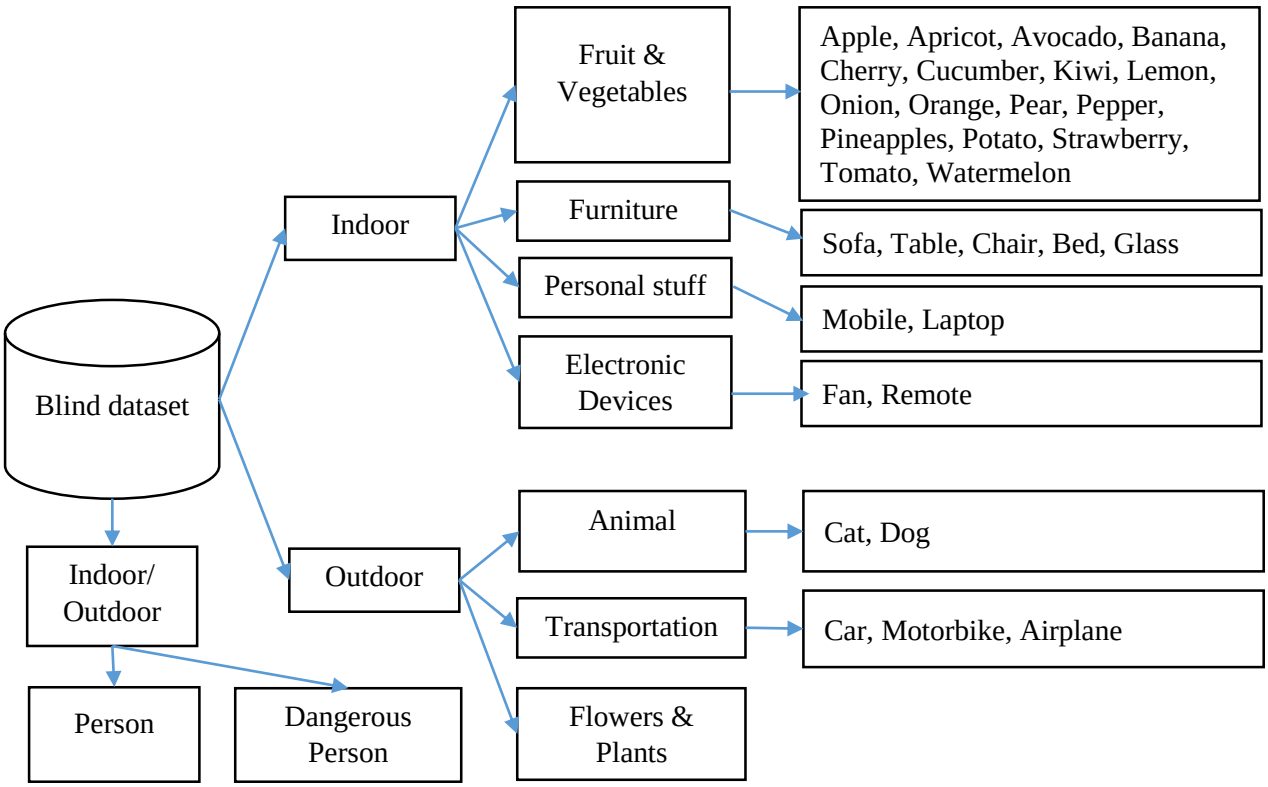


Figure 2. Dataset suggested classes and their corresponding categories.

These classes and sub-classes are chosen based on the most desired stuff that the special-needs people can deal with, use, or even avoid (like dangerous people).

The dataset is collected from different resources, including the following ones:

- Fruit and vegetables dataset: which is one of the most common Kaggle datasets, It's available online[21]. The dataset contains many types of fruits and vegetables, including 22495 images of 33 classes (each image is of size 100*100 and JPEG format). The study uses most of these dataset images.
- Fruit for object detection dataset[22]: This dataset consists of 240 images of 3 classes (orange, apple and banana).
- Natural images dataset[23]: Eight different classes are included in this dataset. A total of 6899 images are involved in this dataset, including airplane, cat, car, dog, flower, fruit, motorbike, and person classes.
- Weapon dataset[24]: This dataset is used only for the "Dangerous people" class.

The final collected dataset includes 13673 images of the indoor classes and 6232 images of the outdoor classes (A total of 19905 images with different sources, dimensions, and formats).

Dataset Split and Augmentation

In any ML or DL system, the dataset is split into training and validation sets. For evaluation purposes, a test set must also be defined.

The collected dataset is split into training (70%), validation (20%) and test (10%) randomly. Another split scenario is suggested by using 30% as a test set.

For the next step, a data augmentation process is applied to all training images. The main purpose of this step is to make new versions of the same samples, making new samples for training and increasing the training set size and improve the learning process by making the model recognise the same sample in different positions, rotations and scaling. The proposed data augmentation includes vertical translation, horizontal translation, reflection, vertical scaling, and horizontal scaling.

2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

GoogleNet Pre-trained Model

GoogleNet is considered one of the CNN architectures that have been used for image classification and recognition. It was first created in the ImageNet Large-Scale Visual Recognition Challenge in 2014 (ILSVRC14)[25].

GoogleNet consists of 22 layers[26], including an essential part called the inception layers. The total number of layers is 27, including the pooling layers.

In DL networks, there are many parameters that control the convolution process, which are stride, padding and kernel size (patch size).

The kernel size refers to the sliding window size of the filter (3x3, 5x5, 7x7, etc.), while the padding is used to take into account the boundary pixels that do not fit into the filter window and need the padding. For example, if the filter is of size 5*5, the padding must be 2 in order to apply convolution on all pixels, making the convolution size the same as the previous layer. The stride defines the number of sliding window steps while moving from one pixel to the next one.

The first layer of GoogleNet is the convolution layer, which has 64 filters with a size of 7*7 and a stride of 2. The next layer is the pooling layer, in which the dimension of the convolved image is reduced from 112*112*64 into 56*56*64 using the max pooling techniques. The third layer is the convolution layer, with 192 filters of size 3*3 and a stride of 2. In the fourth layer, the max pooling layer will reduce the convolution output to 28*28*192.

The next two layers are the inception layers, in which the calculations are performed in parallel. At the output of the first inception layer, the output size is 28*28*256, while the output size of the second layer is 28*28*480. The next five layers are inception layers, followed by a max pooling layer and two inception layers. The output size of these combinations is 7*7*1024. The last four layers are the average pooing layer (acts as a fully-connected layer), the dropout layer with a drop rate of 40%, the linear layer and the Softmax layer in which the classification is performed[27].

The average pooling layer transforms the previous size of 7*7*1024 into one single feature vector of 1*1*1024. Then, some neurons (feature vector samples) of the last feature vector are dropped out to prevent overfitting. The linear layer is used to minimize the size of the feature vector from 1024 to 1000 samples. The Softmax layer is used to compute the probability of all classes, and the class with the maximum probability is considered the final prediction. Table. 1 includes the detailed architecture of GoogleNet.

The GoogleNet architecture was used in many applications through the transfer learning techniques[28-30]. Transfer learning[31] is a technique in which the DL model is trained basically on a specific problem (like image recognition), then it is trained again using different dataset (like cancer detection) so that the knowledge of the first problem is transferred to the next one[31,32]. As long as the two problems are too close, the transfer learning will perform better.

2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Table 1. Common DL networks in the medical domain[25].

Type	Patch Size/ Stride	Output Size	Depth	#1x1	#3x3 Reduce	#3x3	#5x5 Reduce	#5x5	Pool Proj	Params	Ops
Convolution	7x7/2	112x112x64	1							2.7K	34M
Max Pool	3x3/2	56x56x64	0								
Convolution	3x3/1	56x56x192	2		64	192				112K	360M
Max pool	3x3/2	28x28x192	0								
Inception (3a)		28x28x256	2	64	96	128	16	32	32	159K	128M
Inception (3b)		28x28x480	2	128	128	192	32	96	64	380K	304M
Max pool	3x3/2	14x14x480	0								
Inception (4a)		14x14x512	2	192	96	208	16	48	64	364K	73M
Inception (4b)		14x14x512	2	160	112	224	24	64	64	437K	88M
Inception (4c)		14x14x512	2	128	128	256	24	64	64	463K	100M
Inception (4c)		14x14x528	2	112	144	288	32	64	64	580K	119M
Inception (4e)		14x14x832	2	256	160	320	32	128	128	840K	170M
Max pool	3x3/2	7x7x832	0								
Inception (5a)		7x7x832	2	256	160	320	32	128	128	1072K	54M
Inception (5b)		7x7x1024	2	384	192	384	48	128	128	1388K	71M
Avg Pool	7x7/1	1x1x1024	0								
Dropout		1x1x1024	0								
Linear		1x1x1000	1							1000K	1M
Softmax		1x1x1000	0								

MobileNet-V2 Pre-trained Model

MobileNet-V2 is another type of convolutional neural networks[33], consisting of two main parts. The first part is the residual block with a stride of 1, while the other one has a stride of 2, which has the effect of downsizing (minimizing computational time). Two layers exist in each block, the first layer is the 1x1 convolutional layer with the Relu6 activation function, while the second layer performs the depth-wise convolution. The third layer is 1x1 convolution layer without Relu (non-linearity). Fig 3 shows the architecture of MobileNet-V2.

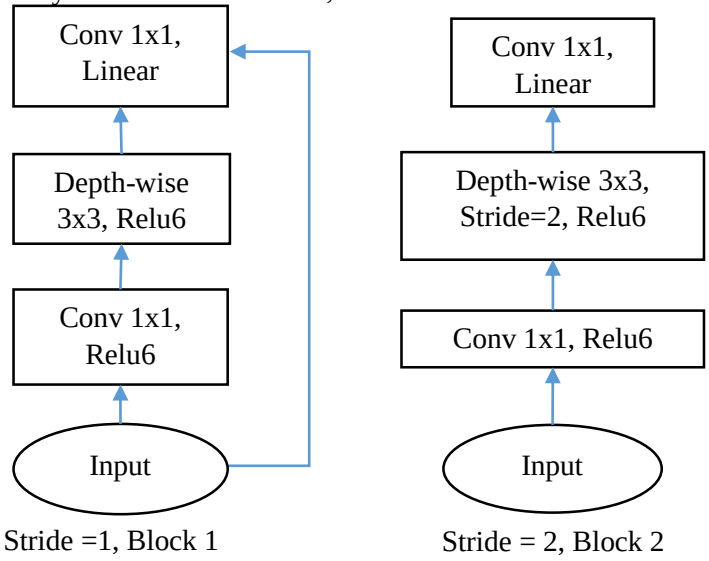


Figure 3. MobileNet-V2 architecture.

2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Results and Discussion

Training Scenarios

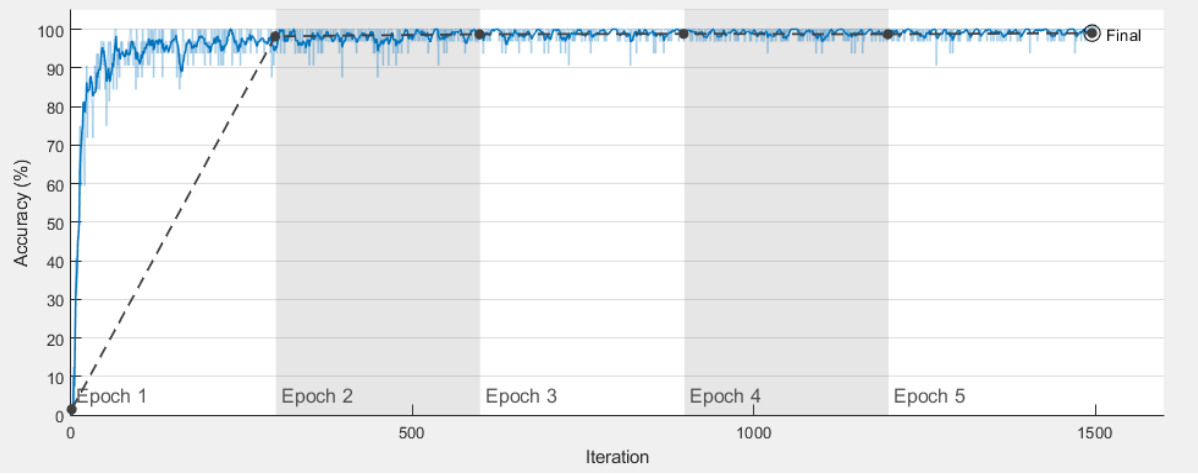
In this step, the GoogleNet and MobileNet-V2 models are trained using the training set. The validation set is also used during the training process. The following scenarios are proposed:

- Indoor category training scenario: In this scenario, the indoor classes are used to train the GoogleNet and MobileNet-V2 pre-trained models, using only 5 epochs and a patch size of 32.
- Outdoor category training scenario: In this scenario, the outdoor classes are used to train the GoogleNet and MobileNet-V2 pre-trained models using the same parameters as the previous scenario.
- All results are repeated using a test set of 30%.

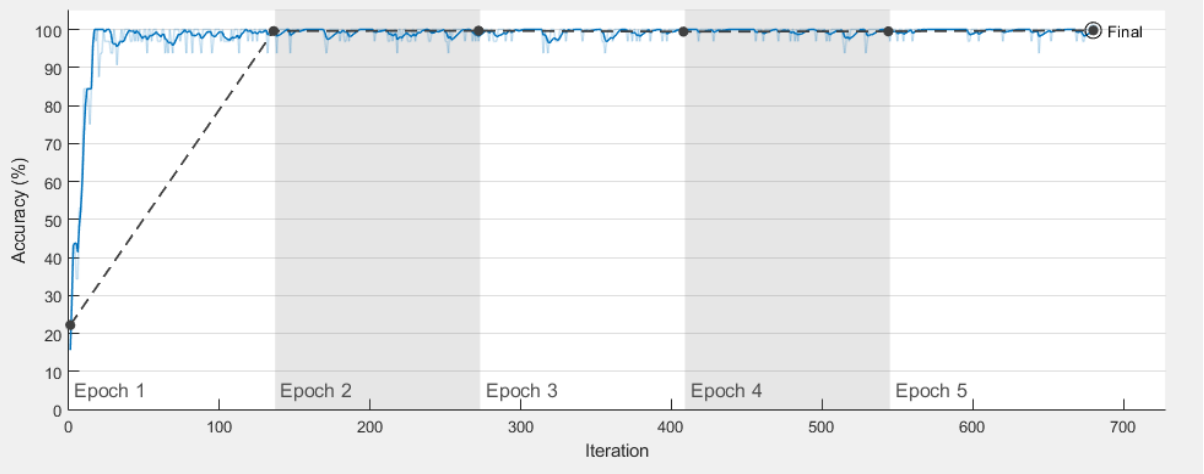
Experimental Results

Experiments were applied using MATLAB and its deep learning toolbox, on an NVIDIA Geforce 750M GPU.

Fig 4 shows the results of training the GoogleNet and MobileNet-V2 models using the indoor and outdoor datasets. Fig 4- a illustrates the training (blue curve) and validation accuracy (black dashed curve) of GoogleNet indoor model, while Fig. 4-b includes the training and validation accuracy of GoogleNet outdoor model. Figs 4-c and 4-d show the training and validation accuracies of the MobileNet-V2 indoor and outdoor models, respectively. Accuracy refers to the number of true classified observations out of all samples. Loss, on the other hand, is another performance metric that predicts the error of a neural network.

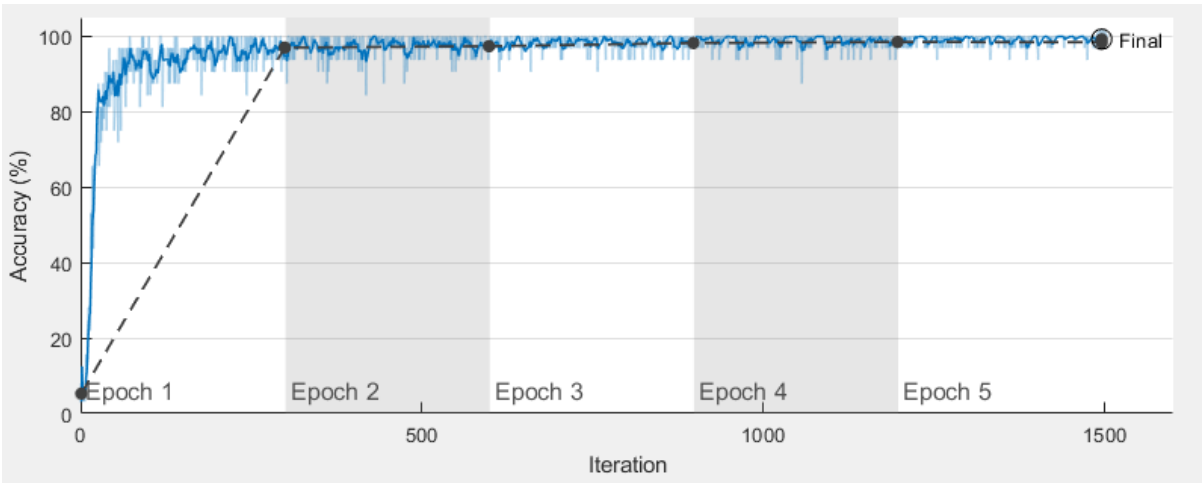


a)

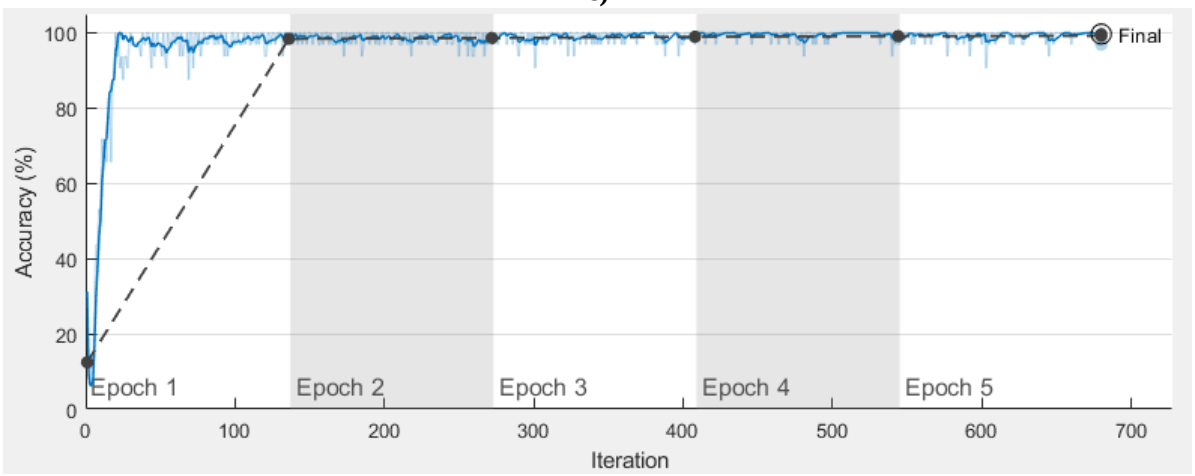


b)

2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal



c)



d)

Figure 4. Accuracy per iterations (a) GooeleNet indoor accuracy, (b) GooeleNet outdoor accuracy, (c) MobileNet-V2 outdoor accuracy, (d) MobileNet-V2 outdoor accuracy.

Fig 4 illustrates that the GoogleNet model gets more than 95% validation accuracy after its first epoch, and MobileNet model gets more than 98% validation accuracy after the first epoch. After 5 epochs, the validation accuracy of indoor GoogleNet and MobileNet-V2 models is 99.01% and 99.27%, respectively. On the other hand, the validation accuracy of the outdoor GoogleNet and MobileNet-V2 models are 99.77% and 99.68%, respectively.

Table 2 shows the detailed results of the indoor-based trained model, including training accuracy, validation accuracy, test accuracy and elapsed training time. The test accuracy is calculated using 10% test set and 30% test set scenarios.

Table 2. Indoor/outdoor training, validation and test accuracies and elapsed time.

Model	Dataset	Training accuracy	Validation accuracy	Test accuracy		Training time
				Test set 10%	Test set 30%	
GoogleNet	Indoor	100%	99.0135%	99.3416%	98.95%	63 min 4 sec
	Outdoor	100%	99.76%	99.1974%	99.36%	28 min 40 sec
MobileNet-V2	Indoor	100%	99.27%	98.68%	98.39%	102 min 59 sec
	Outdoor	100%	99.68%	99.2%	99.57%	43 min 6 sec

Table 2 proves that the indoor and outdoor designed models are robust and accurate. Outdoor training time is higher than outdoor time (in both deep networks) due to the difference in dataset size.

2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Table. 3 ncludes the true positive rate (TPR), false negative rate (FNR), positive predictive rate (PPR) and false discovery rate (FDR) of the validation and test sets for indoor and outdoor categories. Similarly, Table 4 includes the same evaluation metrics using a test set of 30% instead of 10%. GoogleNet achieves a better test accuracy of 0.17% in case of 30% test set for outdoor categories. MobileNet also registered a better performance in case of using 30% test set for outdoor categories, which is similar to GoogleNet result.

TPR is the ratio of correctly classified samples to all samples of the true class. PPR is the ratio of the correctly classified samples to all samples of the predicted class. The FNR is the ratio of incorrectly classified samples to all true class samples. The ratio of incorrectly classified samples per predicted class is called the FDR[34].

Table 3. Indoor/outdoor validation and test performance metrics using 10% test set

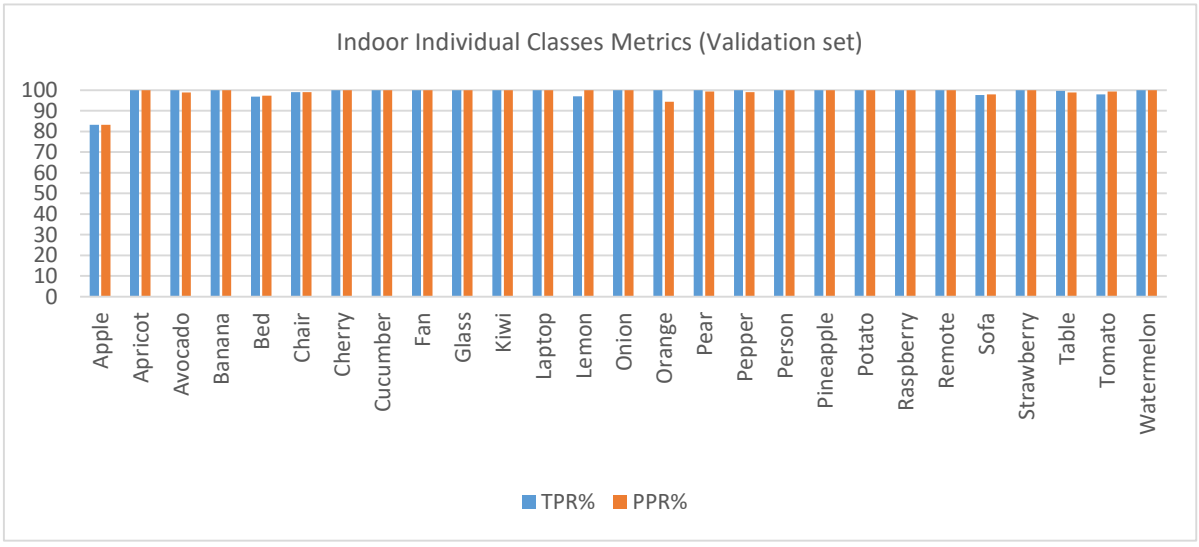
Model	Dataset	Val TPR	Val FNR	Val PPR	Val FDR	Test TPR	Test FNR	Test PPR	Test FDR
GoogleNet	Indoor	98.94%	1.0512%	98.78%	1.2%	95.6%	4.39%	99.29%	0.71%
	Outdoor	99.77%	0.22995	99.753%	0.247%	99.0395%	0.9605%	98.9825%	1.0175%
MobileNet-V2	Indoor	96.36%	3.639%	98.77%	1.225%	98.1397%	1.86%	99.4101%	0.5899%
	Outdoor	99.59%	0.40%	99.655%	0.3448%	99.0455%	0.95%	98.97%	1.029%

Table 4. Indoor/outdoor validation and test performance metrics using 30% test set

Model	Dataset	Val TPR	Val FNR	Val PPR	Val FDR	Test TPR	Test FNR	Test PPR	Test FDR
GoogleNet	Indoor	99.23%	0.764%	97.89%	2.1%	94.95%	5.6%	99.15%	0.84%
	Outdoor	99.39%	0.6%	99.15%	0.85%	99.29%	0.7%	99.54%	0.45%
MobileNet-V2	Indoor	86.68%	13.31%	99%	1%	98.13%	1.87%	98.7%	1.3%
	Outdoor	99.5%	0.5%	99.62%	0.38%	99.53%	0.46%	99.49%	0.5%

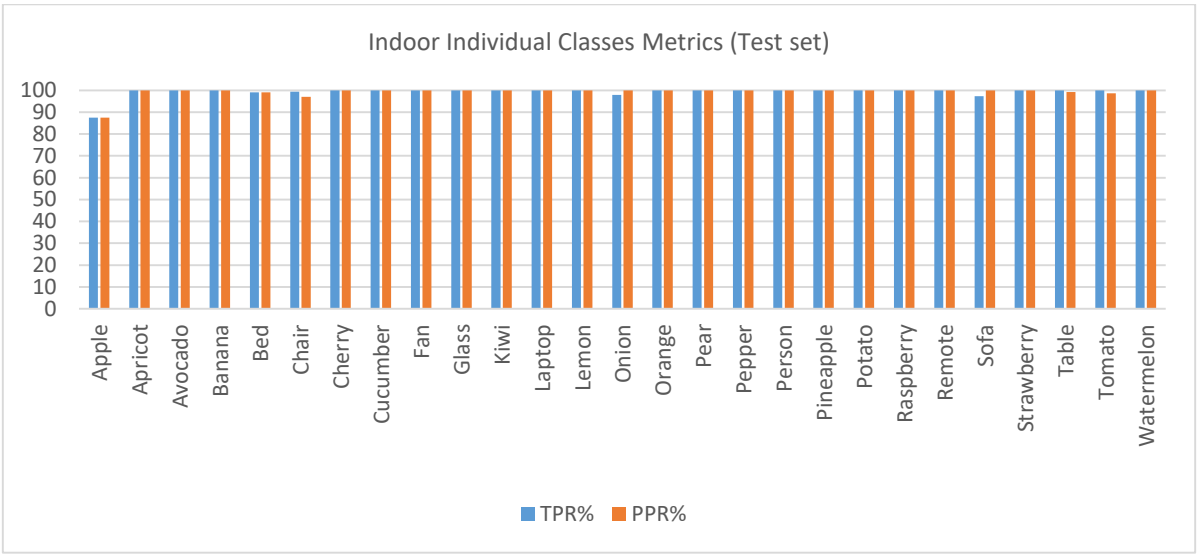
Tables 3 and 4 show that all evaluation metrics for the indoor and outdoor trained models have registered high values in both models. However, the outdoor model metrics are higher by a little bit than the corresponding ones of the indoor model for both the 10% and 30% test set scenarios.

Fig 5 shows the detailed values (TPR and PPR) of the indoor individual classes, while Fig 6 includes the detailed values (TPR and PPR) of the outdoor individual classes.

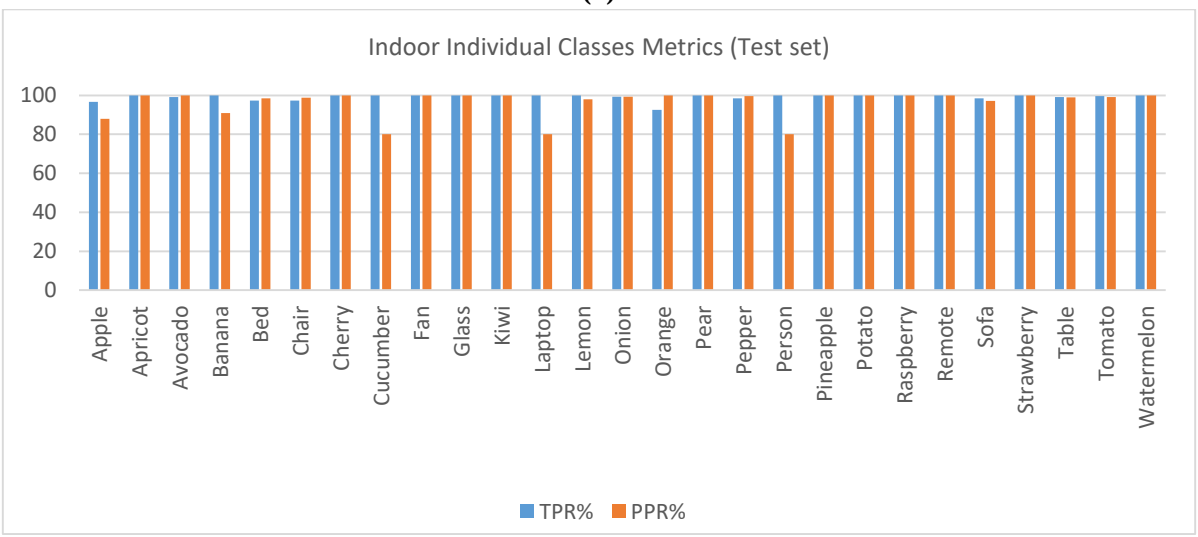


(a)

2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal



(b)



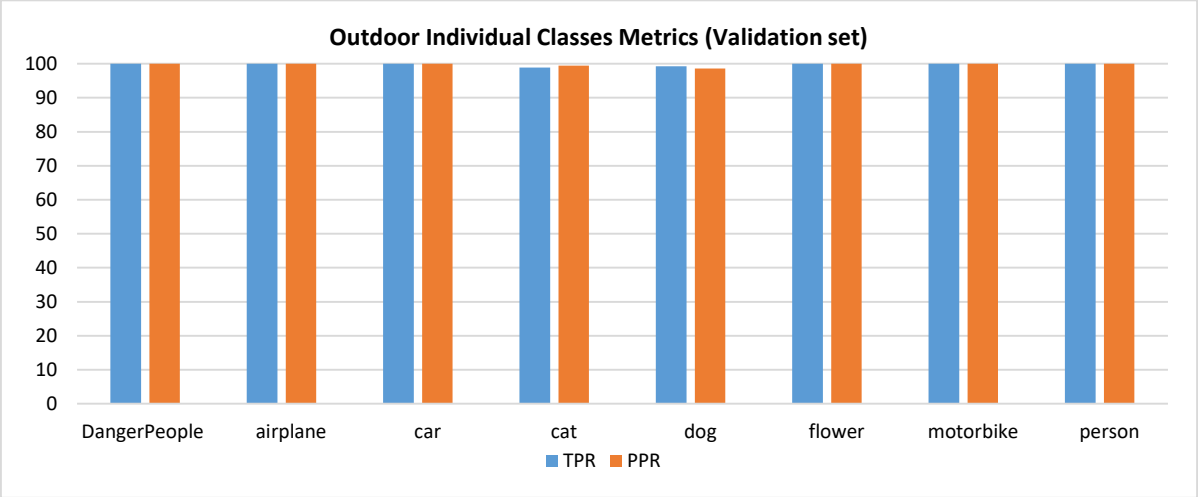
(c)

Figure 5. TPR and PPR of the indoor individual classes (a) Validation set, (b) Test set 10%, (c) Test Set 30%.

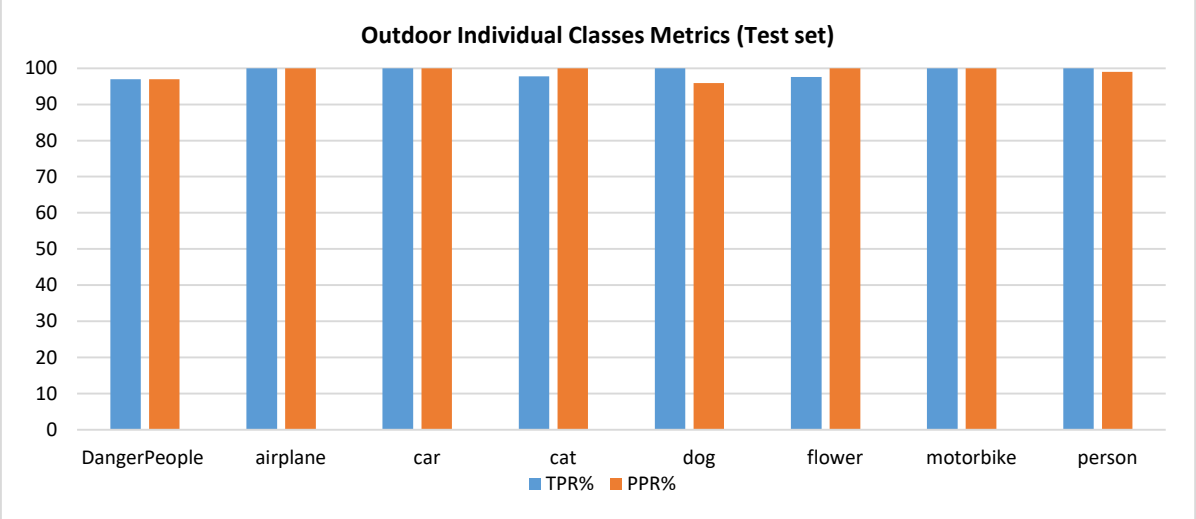
Fig 5 illustrates the individual TPR values of all indoor classes (of the validation set). Using 30% test set gives a similar performance to the "10% test set" case. However, there are some differences since PPR in 10% case was better than in the 30% case. All TPR values have similar values. The lowest TPR value corresponds to the "Apple" class in case of 10% test set, while the TPR value corresponds to the "Apple" and "Orange" classes in case of 30% test set. The confusion matrices of the indoor validation and test sets are illustrated in Fig 7.

One "apple" sample is classified as "lemon" and two others are classified as "tomato", and this is due to the similarity in some samples between these classes and the "apple" class. "bed" class has 7 incorrectly classified samples, 5 of them are classified as "sofa" while two samples are classified as "table". For the test set, as shown in Fig 7-b, three samples of the "sofa" class are misclassified as "chair", while one sample is misclassified as "bed".

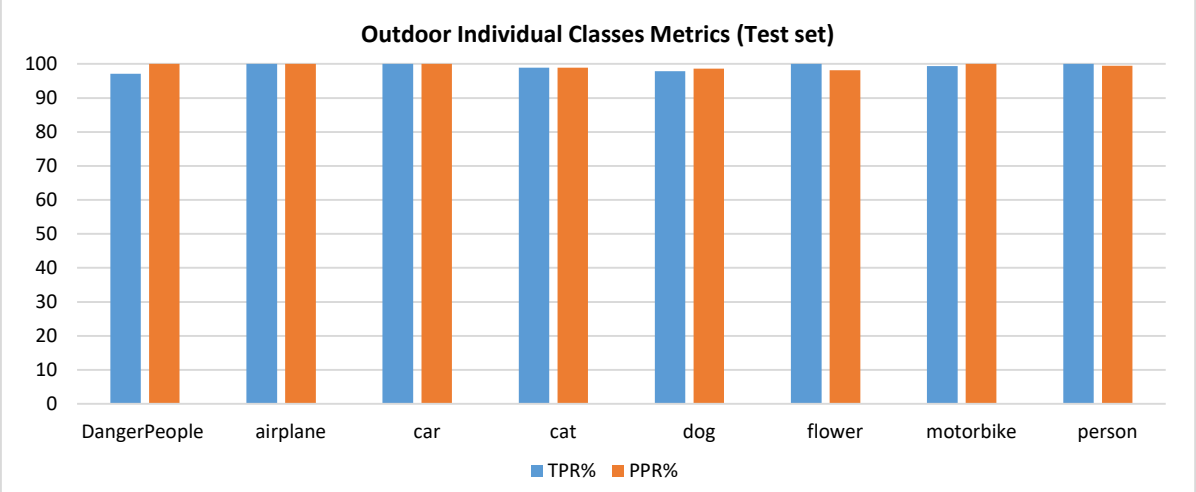
2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal



(a)



(b)



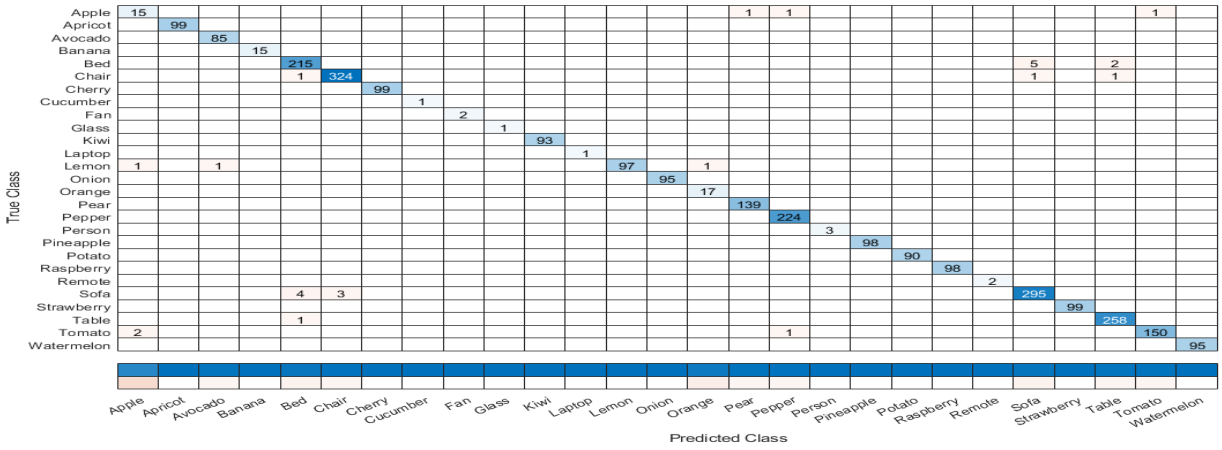
(c)

Figure 6. TPR and PPR of the outdoor individual classes (a) Validation set, (b) Test set 10%, (c) Test Set 30%.

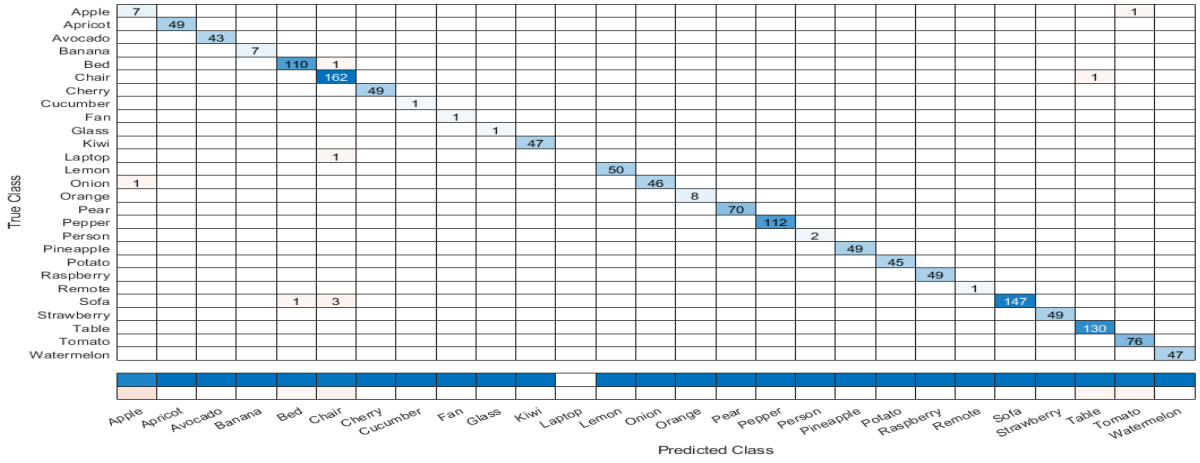
2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

All outdoor TPR and PPR values of the validation and test set have similar values as shown in Fig. 6.

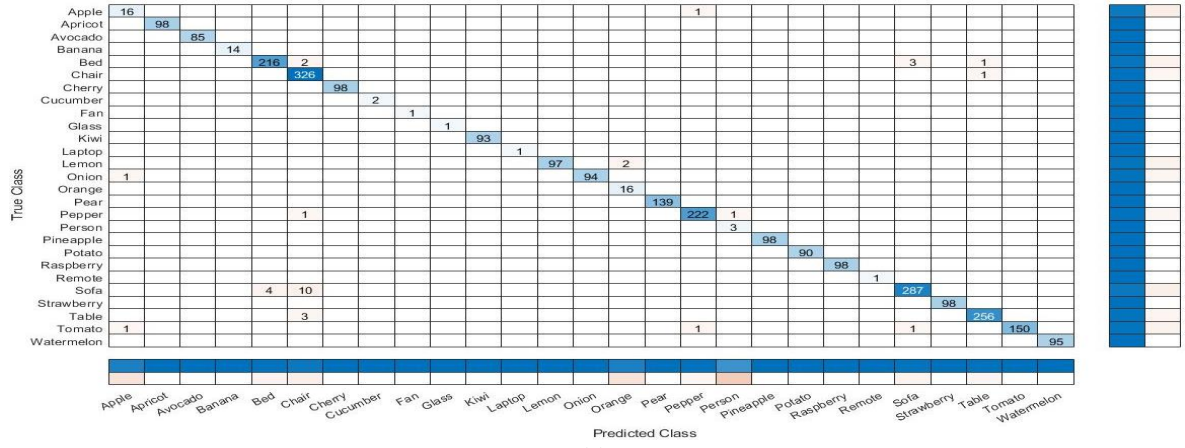
Fig. 8 includes the confusion matrix of the outdoor validation and test sets. Fig 8 shows that two samples of "cat" class are incorrectly classified as "dog", while one sample of class "dog" is classified as "cat", for the validation set. For the test set, there is a misclassification error between "person" and "dangerous Person" classes. All errors are registered between visually similar classes.



(a)



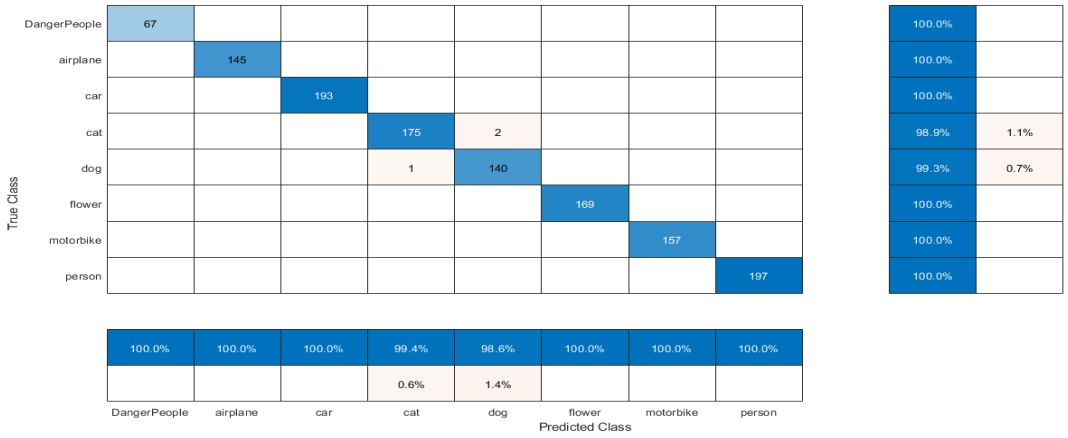
(b)



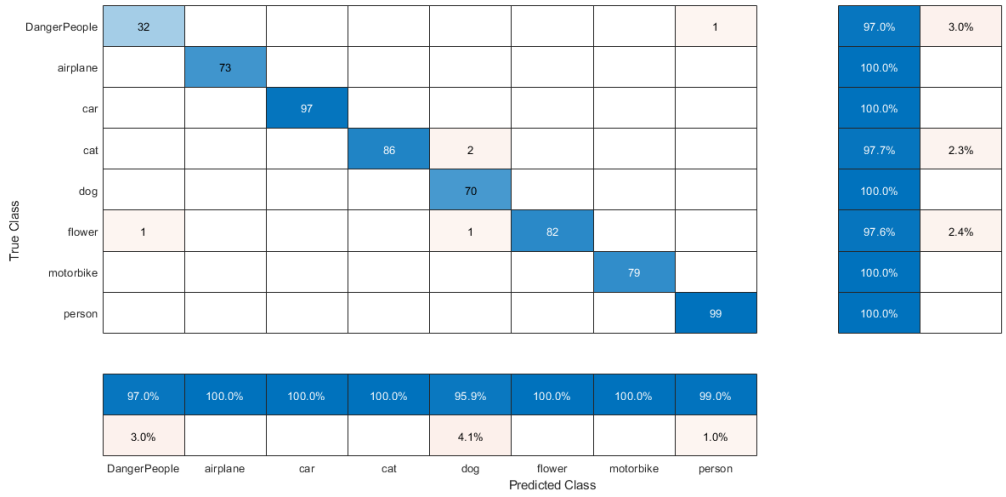
(c)

Figure 7. Confusion matrix of the indoor individual classes (a) Validation set, (b) Test set 10%, (c) 30%.

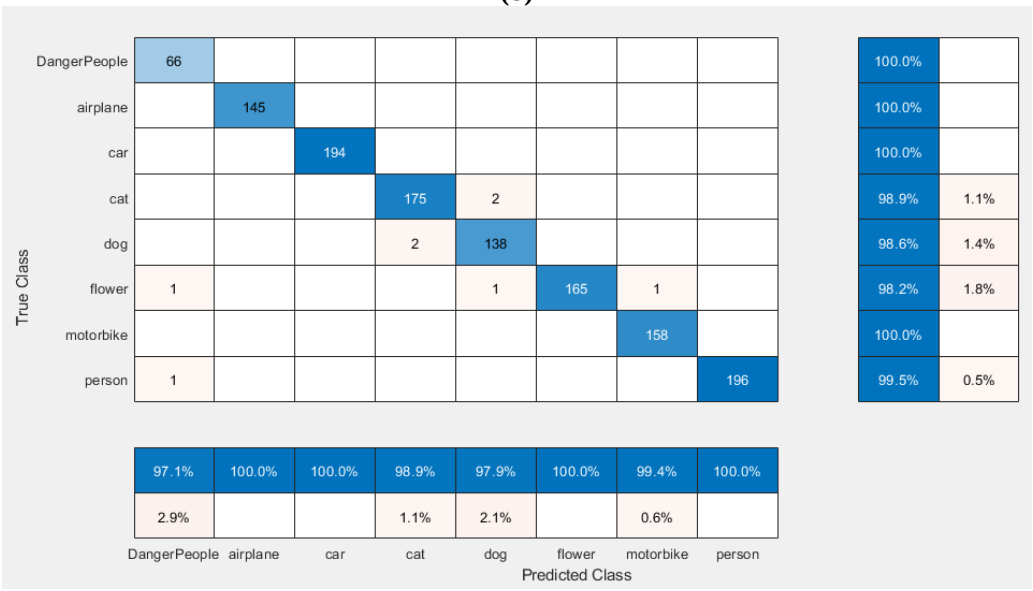
2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal



(a)



(b)



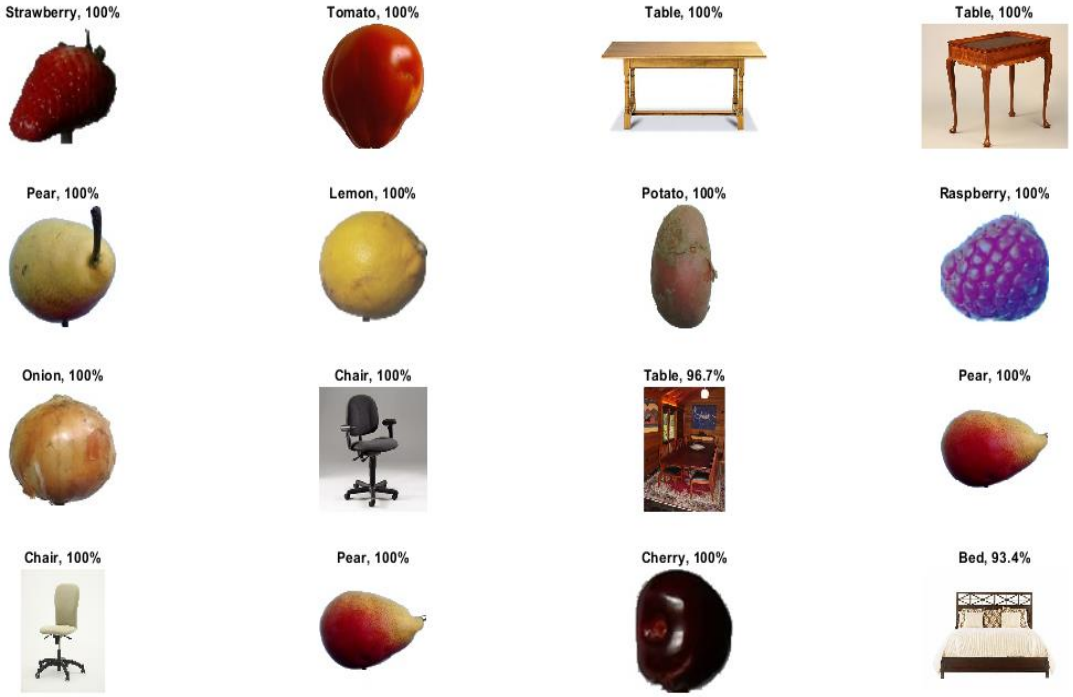
(c)

Figure 8. Confusion matrix of the outdoor individual classes (a) Validation set, (b) Test set 10%, (c) Test set 30%.

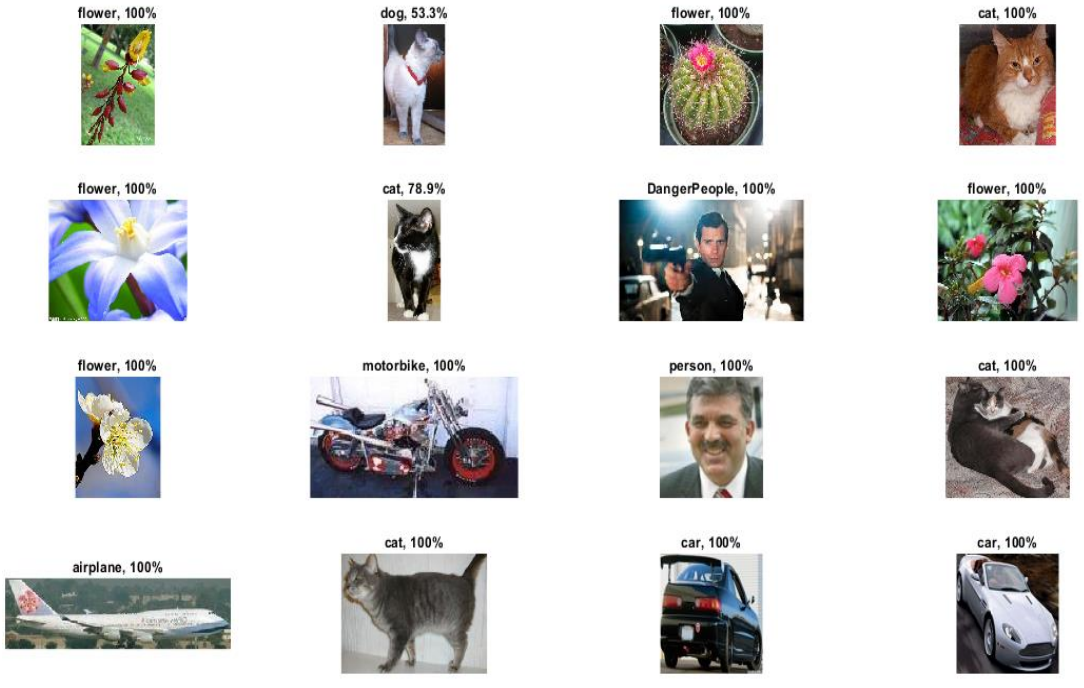
2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Fig 8-c shows that using 30% test set will increase the TPR and PDR values of the outdoor-based trained models compared to the 10% test set case.

Fig 9 includes examples of the indoor and outdoor test samples and their predictions. Each test sample is illustrated with its predicted label and the corresponding score as a percentage (0% to 100%).



(a)



(b)

Figure 9. Examples of test samples and their corresponding predictions (a) Indoor, (b) Outdoor.

2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

The proposed methodology is compared with the most recent similar studies, and the comparative results are listed in Table 5.

Table 5. Results of comparing this study with recent similar researches.

Research	Dataset	Indoor	Outdoor	Methodology	Results
Tapu et al. [10]	60 videos	No	Yes	YOLO CNN	*ACC=90%
Vijaybahadur et al. [11]	Not mentioned	Yes	No	Built-in voice recognition, text-to-speech	Not mentioned
Qureshi et al. [13]	Not mentioned	No	Yes	CNN	ACC=78%
Mone et al. [16]	MsCOCO dataset	No	Yes	MobileNet	Not mentioned
Wang et al. [17]	7 categories each of which contained about 1000 instances	No	Yes	SceneNet and Reinforcement learning	ACC=95.4%
Dhou et al. [20]	Down stairs and hollow pits image dataset	No	Yes	Naïve Bayes, Decision Tree, SVM and k-Nearest Neighbour	ACC=99% to 100%
Current Study	Collected dataset of 36 categories and 19905 images	Yes	Yes	Transfer learning: GoogleNet	ACC=99.34% for indoor dataset, 99.76% for outdoor dataset
				Transfer learning: MobileNet-V2	ACC=99.27% for indoor dataset, 99.68% for outdoor dataset

*ACC: Accuracy

Table 5. proves the robustness and high performance of the current study compared to the previous studies in the same field. The most similar study to the current work is the Dhou et al. study[20], whose attention was only focused on two objects (hollow pits and down stairs); in contrast, the current study takes into account 36 different indoor and outdoor classes.

Conclusion

In the current research, a new dataset of indoor and outdoor items corresponding with the most special-needs people's needs has been created. 36 different classes of indoor and outdoor objects are collected with a total number of 19905 images. This dataset is then split into training, validation and test sets in order to train and evaluate the deep learning models. The transfer learning of GoogleNet and MobileNet-V2 pre-trained networks is used for the training process. Two training scenarios are involved; training the pre-trained models using the indoor dataset and training the pre-trained models using the outdoor dataset.

The test sets of indoor and outdoor datasets are used to evaluate the trained models. Two different test set splits were conducted (10% and 30%). Many performance evaluation metrics are used (training accuracy, validation accuracy, test accuracy, TPR, FNR, PPR and FDR).

Results indicate that the proposed methodology has high performance. The validation accuracies of GoogleNet were 99.01% and 99.76% for indoor/outdoor environments. For MobileNet-V2 model, the validation accuracies were 99.27% and 99.68% for indoor and outdoor environments, respectively. For 10% test set split, the test accuracy of GoogleNet and MobileNet-V2 were 99.34%, 99.197%, 98.68% and 99.2% for indoor and outdoor datasets, respectively. Similarly; for the 30% test set split, the indoor and outdoor accuracies were 98.95%, 99.36%, 98.39% and 99.57% for GoogleNet and MobileNet-V2, respectively.

The FNR and FDR errors are also very small and most of the acquired misclassification errors belong to similar classes (i.e., "bed" and "sofa", "apple" and "lemon", "person" and "dangerous person", etc.).

The limitation of this research is that it didn't take into account all possible classes that special-needs people could need, and future work can use this point as a starting point in the next studies.

2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

The contribution of this study can be concluded with the following:

- Creating a new dataset, including all possible categories that special-needs or ordinary people require (like fruits, vegetables, places, cars, people, dangerous people, etc.)
- Evaluate the trained models by computing not only overall accuracies but also individual's accuracies and errors (FNR and FDR). This can help detect the best and worst classes.

Authors' Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in University of Baghdad.

Authors' Contribution Statement

O. A. J., M. J. A., and Z. H. S. participated in configuring the idea of the paper. O A J collected the dataset and configured the final folders of each category. M J A made the design step, including the architecture of the deep learning networks. Z H S implemented the design along with Omar. All authors participated in the writing part.

References

1. Lv H, Shi S, Gursoy D. A look back and a leap forward: a review and synthesis of big data and artificial intelligence literature in hospitality and tourism. J Hosp Mark Manag. 2022; 31(2): 145-75. https://doi.org/10.1080/19368623.2021.1937434
2. Rozenwald MB, Galitsyna AA, Sapunov GV, Khrameeva EE, Gelfand MS. A machine learning framework for the prediction of chromatin folding in Drosophila using epigenetic features. Peer J Comput Sci. 2020; 6(30). https://doi.org/10.7717/peerj-cs.307
3. Sharma N, Sharma R, Jindal N. Machine learning and deep learning applications-a vision. Glob Trans Proc. 2021; 2(1): 24-8. https://doi.org/10.1016/j.gltp.2021.01.004
4. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021; 8(1): 1-74. https://doi.org/10.1186/s40537-021-00444-8
5. Adeel A, Gogate M, Hussain A. Contextual deep learning-based audio-visual switching for speech enhancement in real-world environments. Inf Fusion. 2020; 1(59): 163-70. https://doi.org/10.1016/j.inffus.2019.08.008
6. Tian H, Chen SC, Shyu ML. Evolutionary programming based deep learning feature selection and network construction for visual data classification. Inf Syst Front. 2020 Oct; 22(5): 1053-66. https://doi.org/10.1007/s10796-020-10023-6
7. Lee SB, Gui X, Manquen M, Hamilton ER. Use of training, validation, and test sets for developing automated classifiers in quantitative ethnography. Int Conf Quant Ethn. 2019; 117-127. https://doi.org/10.1007/978-3-030-33232-7_10
8. Vishwakarma M, Singh HP, Kumar N, Arora M. The Need of Smart Guidance Systems for Blind People in the World. Proc Int Conf Big Data Mach Learn App. 2021; 191-195. https://doi.org/10.1007/978-981-15-8377-3_17
9. Durgadevi S, Thirupurasundari K, Komathi C, Balaji SM. Smart machine learning system for blind assistance. Int Conf Power Energy Control Trans Syst. IEEE. 2020; 1-4. https://doi.org/10.1109/ICPECTS49113.2020.9337031
10. Tapu R, Mocanu B, Zaharia T. DEEP-SEE: Joint object detection, tracking and recognition with application to visually impaired navigational assistance. Sensors. 2017; 17(11): 2473. https://doi.org/10.3390/s17112473
11. Yadav AV, Verma SS, Singh DD. Virtual Assistant for blind people. Int J Adv Sci Res Eng Trends. 2021; 6(5):156-159. http://ijasret.com/VolumeArticles/FullTextPDF/83136.VIRTUAL_ASSISTANT_FOR_BLIND_PEOPLE.pdf
12. Ephzibah EP. Assisting Blind People Using Machine Learning Algorithms. Tur. Co. Mat. 2021; 12(8): 3162-70. https://doi.org/10.17762/turcomat.v12i8.4161
13. Qureshi TA, Rajbhar M, Pisat Y, Bhosale V. AI Based App for Blind People. Int Res J Eng Technol. 2021; 8(03): 2883-7. https://doi.org/10.1177/02646196221131746

14. Mamun SA, Daud ME, Mahmud M, Kaiser MS, Rossi AL. ALO: AI for least observed people. Int Conf Appl Intell Inform. Springer. 2021; 306-317. https://www.springerprofessional.de/en/alo-ai-for-least-observed-people/19396306
15. Chaurasia MA, Rasool S, Afroze M, Jalal SA, Zareen R, Fatima U, et al. Automated Navigation System with Indoor Assistance for Blind. In Contactless Healthcare Facilitation and Commodity Delivery Management During COVID 19 Pandemic. Springer, Singapore. 2022; 119-128. https://doi.org/10.1007/978-981-16-5411-4_10
16. Mone S, Salunke N, Jadhav O, Barge A, Magar N. Machine Learning Based Computer Vision Application for Visually Disabled People. Int J Sci Res Comput Sci Eng Inf Technol. 2021; 7(3): 488-494. https://doi.org/10.32628/CSEIT2173130
17. Wang K, Chen CM, Hossain MS, Muhammad G, Kumar S, Kumari S. Transfer reinforcement learning-based road object detection in next generation IoT domain. Comput Netw. 2021; 193:1-12. https://doi.org/10.1016/j.comnet.2021.108078
18. Bouteraa Y. Design and Development of a Wearable Assistive Device Integrating a Fuzzy Decision Support System for Blind and Visually Impaired People. Micromachines. 2021; 12(9): 1082. https://doi.org/10.3390/mi12091082
19. Periša M, Peraković D, Cvitić I, Krstić M. Innovative ecosystem for informing visual impaired person in smart shopping environment: IoT Shop. Wirel Netw. 2022; 28(1): 469-79. https://doi.org/10.1007/s11276-021-02591-5
20. Dhou S, Alnabulsi A, Al-Ali AR, Arshi M, Darwish F, Almaazmi S, et al. An IoT machine learning-based mobile sensors unit for visually impaired people. Sensors. 2022; 22(14): 5202. https://doi.org/10.3390/s22145202
21. Zhang E, fruit recognition dataset. Kaggle. 2022. [Online]. https://www.kaggle.com/datasets/sshikamaru/fruit-recognition .
22. Buyukkinaci M., fruit images for object detection dataset. Kaggle. 2018. https://www.kaggle.com/datasets/mbkinaci/fruit-images-for-object-detection .
23. Roy P, Ghosh S, Bhattacharya S, Pal U. Natural images dataset. Kaggle, 2018. https://www.kaggle.com/datasets/prasunroy/natural-images .
24. Annamraju. Weapon detection dataset. Kaggle. 2019. https://www.kaggle.com/datasets/abhishek4273/gun-detection-dataset .
25. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. Proc. IEEE conf Comput Vis Pattern Recognit. 2015: 1-9. https://doi.org/10.48550/arXiv.1409.4842
26. Hub T. Convolutional neural network architectures. Principles and Labs for Deep Learning. 2021: 6: 201. https://doi.org/10.1016/C2020-0-03408-0
27. AL-Huseiny MS, Sajit AS. Transfer learning with GoogLeNet for detection of lung cancer. Indones J Electr Eng. 2021; 22(2): 1078-86. http://doi.org/10.11591/ijeecs.v22.i2.pp1078-1086
28. Sharma S, Kumar H. Detection and classification of plant diseases by Alexnet and GoogleNet deep learning architecture. Int J Sci Res Eng Trends. 2022; 8(1): 218-23. https://ijsret.com/wp-content/uploads/2022/01/IJSRET_V8_issue1_120.pdf
29. Zakaria N, Mohamed F, Abdelghani R, Sundaraj K. VGG16, ResNet-50, and GoogLeNet Deep Learning Architecture for Breathing Sound Classification: A Comparative Study. Int Conf Artif Intell Cyber Secur Syst. IEEE. 2021:1-6. https://doi.org/10.1109/AI-CSP52968.2021.9671124
30. Mayya A, Khozama S. A. Novel Medical Support Deep Learning Fusion Model for the Diagnosis of COVID-19. Int Conf Adv Trends Multidiscip Res Innov. IEEE. 2020: 1-6. https://doi.org/10.1109/ICATMRI51801.2020.9398317
31. Abdullah TH, Alizadeh F, Abdullah BH. COVID-19 Diagnosis System using SimpNet Deep Model. Baghdad Sci J. 2022; 19(5): 1078-1089. DOI: http://dx.doi.org/10.21123/bsj.2022.19.4.ID0000.
32. Zaki SM, Jaber MM, Kashmoola MA. Diagnosing COVID-19 Infection in Chest X-Ray Images Using Neural Network. Baghdad Sci J. 2022; 19(6): 1356-1361. DOI: https://dx.doi.org/10.21123/bsj.2022.5965.
33. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. Proc IEEE conf Comput Vis Pattern Recognit. 2018: 4510-4520. https://doi.org/10.48550/arXiv.1801.04381
34. Feigenbaum J J. A machine learning approach to census record linking. Harvard University. 2016. https://ranabr.people.stanford.edu/sites/g/files/sbiybj26066/files/media/file/machine_learning_approach.pdf

2023, 20(6 Suppl.): 2540-2558
https://dx.doi.org/10.21123/bsj.2023.8177
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

تصنيف الصور القائم باستخدام التعلم العميق لتطبيقات كشف النماذج داخل وخارج المنزل

عمر عبد اللطيف جاسم[1]، محمد جواد عبد[1] ، زينه هادي سعيد[2]

[1] قسم هندسة تقنيات الاجهزة الطبية, كلية الحكمة الجامعة, بغداد, العراق
[2] قسم تقنيات المختبرات الطبية, المعهد الطبي التقني-المنصور, الجامعة التقنية الوسطى, بغداد, العراق

الخلاصة

مع التطور السريع في تصميم الأجهزة الذكية، أصبحت حياة الناس أسهل خصوصاً أولئك الذين يعانون من فقدان البصر أو العمى. الإنجازات الجديدة في مجال تعلم الآلة والتعلم العميق سمحت لفاقدي البصر بالتعرف على البيئة المحيطة بهم وتمييزها. في الدراسة الحالية، نقوم باستخدام الفعالية والأداء العالي الذي تتمتع به أنظمة التعلم العميق لبناء نظام تصنيف الصور في كلا البيئتين الداخلية والخارجية. تبدأ الطريقة المقترحة بإنشاء مجموعتي بيانات داخلية وخارجية من عدة مصادر بيانات مختلفة. في الخطوة التالية، يتم تقسيم مجموعة البيانات المجمعة إلى مجموعات تدريب وتحقق واختبار. يتم استخدام نموذجي التعلم العميق المدربين مسبقاً المسميين GoogleNet وMobileNet-V2 من التدريب باستخدام مجموعتي البيانات الداخلية والخارجية وينتج عن ذلك نموذجان مدربان. يتم استخدام مجموعات بيانات الاختبار من أجل اختبار النماذج المدربة باستخدام معاملات قياس الأداء (الدقة، معدل القبول الصحيح، معدل الرفض الخاطئ، معدل التخمين الصحيح، ومعدل الاكتشاف الخاطئ). بالنسبة لنموذج GoogleNet تشير النتائج إلى الأداء العالي للأنظمة المدربة حيث تم التوصل لدقات اختبار 99.34% و99.76% لكل من مجموعتي البيانات الداخلية والخارجية على التوالي. أما فيما يخص نموذج MobileNet فقد تم التوصل لدقات 99.27% و 99.68% لكل من مجموعي بيانات الداخلية والخارجية على التوالي. تمت مقارنة الطريقة المقترحة مع الطرق المماثلة في الدراسات السابقة في مجال تصنيف الصور في أنظمة رعاية فاقدي البصر، حيث أظهرت تفوق الطريقة المقترحة من قبلنا.

الكلمات المفتاحية: التعلم العميق، نموذج التعلم المدرب مسبقاً واسمه GoogleNet، تصنيف الصور، المشاهد داخل وخارج المنزل، نقل التعلم.