Journal of King Saud University – Computer and Information Sciences 34 (2022) 578–588

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

# Convolutional neural networks based potholes detection using thermal imaging

Aparna [a], Yukti Bhatia [a], Rachna Rai [a], Varun Gupta [a,*], Naveen Aggarwal [b], Aparna Akula [c]

[a] Department of Computer Science and Engineering, Chandigarh College of Engineering and Technology, India
[b] UIET, Panjab University, Chandigarh, India
[c] Computational Instrumentation, CSIR-Central Scientific Instruments Organisation, Chandigarh, India

ARTICLE INFO

*Article history:*
Received 4 December 2018
Revised 2 February 2019
Accepted 3 February 2019
Available online 5 February 2019

*Keywords:*
Pothole detection
Thermal imaging
Convolutional neural networks
Residual networks
Deep learning

ABSTRACT

The presence of potholes on the roads is one of the major causes of road accidents as well as wear and tear of vehicles. In order to solve this problem, various techniques have been implemented ranging from manual reporting to authorities to the use of vibration-based sensors to 3D reconstruction using laser imaging. But all these techniques have some drawbacks such as the high setup cost, risk while detection or no provision for night vision. Therefore, the objective of this work is to analyze the feasibility and accuracy of thermal imaging in the field of pothole detection. After collecting a suitable amount of data containing the images of potholes under various conditions and weather, and implementing augmentation techniques on the data, convolutional neural networks approach of deep learning has been adopted, that is a new approach in this problem domain using thermal imaging. Also, a comparison between the self-built convolutional neural model and some of the pre-rained models has been done. The results show that images were correctly identified with the best accuracy of 97.08% using one of the pre-trained convolutional neural networks based residual network models. The results of this work will be helpful in guiding the future researches in this novel application of thermal imaging in pothole detection field.



## 1. Introduction

Poor road conditions due to potholes have been a major cause of road accidents and damage to vehicles. Recently, with an increase in vehicular traffic and pollution, the roads are getting filled with big and small potholes in almost every city in the country. Potholes took a deadly toll in 2017, claiming almost 10 lives daily with annual fatalities in the country adding up to 3597—a more than 50% rise over the toll for 2016 (Dash, 2018). This is a major problem in many developed countries also. A large amount of money is sanctioned by the authorities to fill the potholes but due to inefficient detection of these potholes, they are not dealt with on time and thus, become the cause of accidents and other mishaps.

Detecting potholes manually is a labour-intensive and time-consuming task. Many techniques have been implemented to detect potholes like vibration-based methods, 3D reconstruction-based methods, and vision-based methods. But each of these techniques has some limitations. Thus, we have tried to address this problem of potholes using a novel technique of thermal imaging that uses the infrared rays emitted by objects to form images based on the temperature differences.

### 1.1. Thermal imaging v/s optical imaging

All objects emit infrared energy (heat) as a function of their temperature. The infrared energy emitted by an object is known as its heat signature. In general, the hotter an object is, the more radiation it emits. A thermal imager (also known as a thermal camera) is essentially a heat sensor that is capable of detecting tiny differences in temperature. The device collects the infrared radiations from objects in the scene and creates an electronic image based on information about the temperature differences. Since objects are rarely of the same temperature as of the other objects around them, a thermal camera can detect them, and they will appear as distinct in a thermal image (Negied, 2014). It is different from optical imaging in terms of image formation. Optical imaging forms

* Corresponding author.
*E-mail address:* varungupta@ccet.ac.in (V. Gupta).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2019.02.004
1319-1578/© 2019 The Author(s). Published by Elsevier B.V. on behalf of King Saud University.

image with the help of light whereas thermal imaging forms images with the heat emitted by the objects. This is the main advantage of thermal imaging in the detection of potholes as it can be used in areas that are not well lit or having weather conditions like fog and rain. Thermal imaging is also cheaper than other night vision techniques such as laser-based image reconstruction.

Through this work, the technology of thermal imaging has been used to capture images for preparing dataset of potholes and then convolutional neural networks have been used for prediction of the presence of potholes in thermal images.

### 1.2. Motivation

Potholes on roads pose a great danger to lives and this is why it is an important issue to be addressed. Since the root cause of pothole formation is water logging and due to the presence of water in cracks and crevices, the temperature of potholes is normally expected to be less than the surrounding road. Thus, it is easy to differentiate between pothole and non-pothole using the thermal imaging technique.

Also, normal vision cameras can miss potholes at night but since thermal cameras detect heat and not light, this disadvantage is overcome. There are many other advantages of using thermal cameras such as:

- High response time as compared to vibration-based sensors. There is no need to pass through a pothole to sense the presence of the pothole.
- Less energy consumption – These can be charged using the car battery.
- Cheaper than the existing night-time detection techniques like laser-based techniques.
- Not affected by visual occlusions and lighting conditions, thus, they can be used during day as well as night.
- Image processing techniques used in vision-based cameras can be used for thermal imaging too.

### 1.3. Contribution

Through this work, a new benchmark in the domain of pothole detection has been set as this work uses thermal imaging for the collection of data and uses a convolutional neural network for the training, which is also new to this field with such a dataset. Another contribution of this work is the dataset. The data has been generated by manual capturing of images of potholes with the help of a thermal camera. There is no pre-existing dataset of thermal images of potholes. Thus, this can also contribute future researches in this field.

## 2. Related work

Various techniques are being used for pothole detection like manual notification using mobile applications, computer vision-based techniques, sensor-based techniques and many more. The simplest technique to solve this problem involves the use of a mobile application such that whenever a user comes across any pothole on road, he or she clicks the picture of the pothole and sends it along with the location to the concerned authority so that action can be taken. This technique, although simple, requires human intervention and also depends on the voluntariness of the people. Thus, this technique may not prove to be of great use given the amount of money spent in developing the application and promoting it among people. Thus, automated pothole detection techniques are more useful.

Existing methods for pothole detection can be classified broadly into 3 types as (Kim and Ryu, 2014): Vibration-based methods (Yu and Yu, 2006; De Zoysa et al., 2007; Erikson et al., 2008; Mednis et al., 2001), 3D reconstruction-based methods (Wang, 2004; Chang et al., 2005; Hou et al., 2007; Li et al., 2009; Staniek, 2013; Joubert et al., 2011; Moazzam et al., 2013) and Vision-based methods (Koch and Brilakis, 2011; Jog et al., 2012; Lokeshwor et al., 2013; Koch et al., 2013; Buza et al., 2013; Lokeshwor et al., 2014).

Table 1 summarizes and compares various pothole detection methods. However, techniques mentioned in Table 1 have some problems or drawbacks such as (Kim and Ryu, 2014):

1) High cost of setup and equipment.
2) Low accuracy due to the problem in detecting at night or foggy weather.
3) Risk in passing through potholes.
4) Complex algorithms involved.

N. Hoang (Hoang, 2018) proposed an artificial intelligence model for pothole detection and trained it using two machine learning algorithms including the least squares support vector machine (LS-SVM) and the artificial neural network (ANN). This work achieved the classification accuracy rate of approximately 89% using the LS-SVM algorithm and roughly 86% using ANN.

K. An *et al.* (2018) experimented with several pre-trained models for pothole detection. They achieved the classification accuracy rate of 97% using coloured images and 97.5% using grayscale images.

S. Ryu *et al.* (2015) designed a pothole detection system to collect road images through an optical device mounted on a vehicle and then proposed an algorithm to detect potholes from the collected data. According to the proposed algorithm; first, a histogram and the closing operation of a morphology filter are used for extracting dark regions for the pothole. Next, candidate regions of a pothole are extracted using various features, such as size and compactness. Finally, a decision is made whether candidate regions are potholes or not, by comparing pothole and background features. Using this technique, they achieved the classification accuracy rate of 73.5%.

A. Akagic *et al.* (2017) proposed an unsupervised vision-based method for pothole detection. According to the proposed method, first, the pothole areas are extracted from the RGB colour space and then, image segmentation is performed on them. Afterwards, the search is done in the extracted region of interest (ROI) only. Their method is suitable as a pre-processing step for other supervised methods. The effectiveness of their method depends on the extraction of ROI accuracy and they achieved an accuracy of 82%.

Some other related works include the work done by Chun-Yu Yang et al. (2019) where they considered the boundary value problems for differential equations in fractal heat transfer. In this work, the exact solutions of the non-differentiable type were obtained by using the local fractional differential transform method. Other similar work was done by Shui-Hua Wang et al. (2017). They developed a method using extreme learning machine and a combination of statistical measure and fractal dimension for detection of breast disease 'ductal carcinoma in situ (DCIS)' based on breast thermography, which can provide earlier alerts.

## 3. Proposed scheme

The main objective of this work is to develop a system that can detect potholes from thermal images. This pothole detection could be done in real time. For this, a deep learning-based approach has been used. A convolutional neural network (CNN)-based model has been designed that takes as input- thermal images of potholes and

**Table 1**
Different pothole detection methods.

| | Vision-based | Vibration-based | Laser-based | Stereo imaging |
|---|---|---|---|---|
| Device used | Camera | Accelerometer | Laser | Cameras |
| Technology used | 2D imaging | Force and rotation and orientation | 3D reconstruction of the image using light reflection | 3D reconstruction using multiple cameras |
| Response time | High | Low | Low | High |
| Sensing time | While approaching the pothole | While going through a pothole | While approaching the pothole | While approaching the pothole |
| Processing | Complex image processing algorithms | Readings are directly used | Collection of 3D point cloud with their elevations. | A complex process of 3D image construction by combining image from different camera perceptions |
| Cost | High because of delicate parts like lens | Low | High | High |
| Characterization of pothole | Based on size | Based on vibrations | Based on 3D image constructed | Based on 3D image constructed |
| Detection at night time | Difficult due to poor lighting | Can detect | Can detect | Difficult due to poor lighting |
| Accuracy | Depends on the algorithm used | High | High | Depends on the alignment of cameras and algorithms used |

non-pothole roads. After training the model on this data, the model predicts if the input image is of a pothole or non-pothole. The convolutional neural networks have been used in various fields such as radiology (Yamashita et al., 2018), for tasks like classification, object detection, segmentation etc. Further, we have used pre-trained neural network models based on residual networks for obtaining better results for the given task. However, we have not used the pre-trained models just off-the-shelf. First, we fine-tuned these models for the given problem by training the model on the pre-computed weights for few epochs. Then, we discarded precomputed weights of last layers and trained the model using best practices such as cyclic differential learning rates and test time data augmentation. The results obtained from these models are then compared. Also, a comparison between the proposed work and the previously existing works has been done to analyse if the proposed model based on convolutional neural networks on thermal images is efficient and feasible.

### 3.1. Self-Built CNN model

First, we developed a convolutional neural network model from scratch. The architecture of the self-built CNN model is discussed below.

1) We developed a sequential model, where layers are connected sequentially to each other. In this model, the input is given to the batch normalization layer where it is normalized to help the model effectively learn the parameters.
2) The normalized input is passed to a series of 2-dimensional convolution layers with 3x3 kernel and ReLU activations. These layers help in the extraction of certain features from the image.
3) Each convolution layer is followed by a max pooling layer which helps in reducing the dimension of input further.
4) The output after max pooling is normalized using batch normalization layer and passed to the next block.
5) After passing through the convolution layers, the output is passed through a global average pooling which reduces the size of output through averaging a neighbourhood.
6) Finally, the output of the previous layer enters as an input to the dense layer with one neuron that finally classifies the input as 0 or 1 using sigmoid activation.
7) The model uses binary cross-entropy as the loss function which is a logarithmic loss function.
8) Adam optimizer with various parameters like learning rate, decay etc. has been used for optimization.

Fig. 1 presents the architecture of the proposed self-built CNN model developed from scratch.

### 3.2. Transfer learning from pre-trained residual network models

Transfer learning is a method that allows us to build accurate models in a time-saving way. Instead of starting the learning process from scratch, we choose a model that has already been trained in a similar and a much larger dataset to solve a similar problem. We import a pre-trained model in accordance to our problem domain and size of our dataset and then fine-tune the model to perform the given task. S. Rajaraman *et al.* (2018) and C. V. Roberts (2018) have determined in their works how fine-tuning pre-trained models on a large dataset results in improved performance. Thus, we have experimented with various pre-trained ResNet models in the proposed work.

It has been seen that as the depth of the network increases, the error rate also increases, but sometimes it is required to create a deep network. The residual networks (ResNet) solve this problem with the help of adding residual layers above identity mappings. It is easy to reduce the residual output to a minimum than the output of non-linear layers (He et al., 2016). Each ResNet block is either 2 layers deep (used in small networks like ResNet 18, 34) or 3 layers deep (ResNet 50, 101, 152) (He et al., 2016).

We have used different models of ResNet and with different image sizes. ResNet model expects an image of square size as input. We have tried to employ the best practices for training the models. We have used cyclic learning rates that change their values cyclically between epochs and differential learning rates that change their values according to layers such that the training rate varies layer-wise as initial layers usually represent primitive features and inner layers represent high-level features. This helps in improving the accuracy greatly and reducing overfitting if any. For the purpose, we have used the Fastai library on top of Pytorch, a deep learning library in Python. This helped us in achieving the best results for the given problem.

## 4. Methodology

In this section, we present the methodology used for the proposed work. The objectives of the proposed work are as follows:

1) To find efficient and accurate CNN models for pothole detection using thermal imaging technique.
2) To evaluate the proposed technique with respect to the existing techniques.

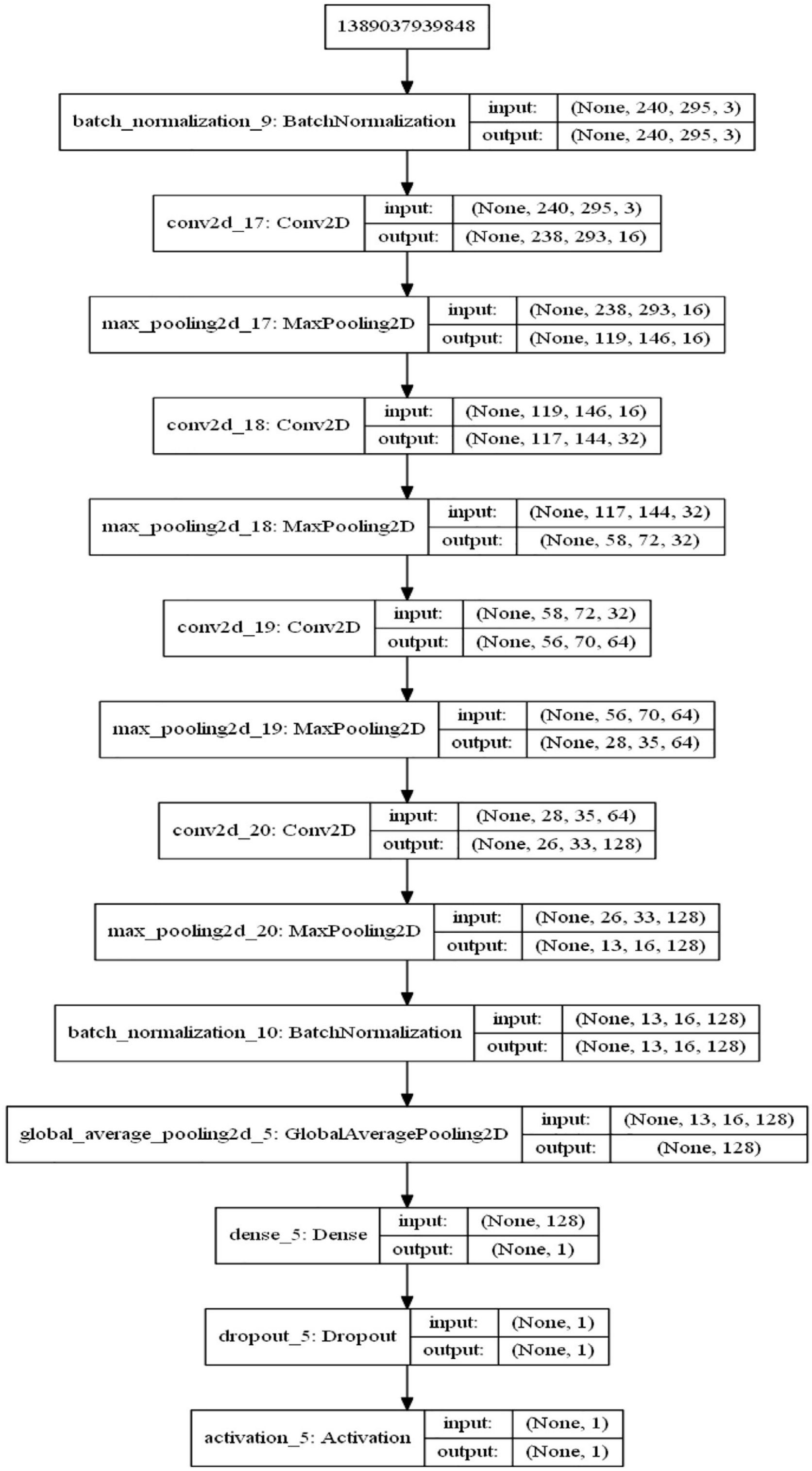



**Fig. 1.** Self-built CNN model architecture.

In order to carry out the proposed work, we have chosen the deep learning especially convolutional neural networks on thermal images dataset since it is state-of-the-art technology that can be used to solve most of the problems be it classification or object detection in almost every domain. Also, this technology has not been yet applied for solving potholes problem using thermal imaging. Thus, our plan was to collect data containing thermal images of roads with and without potholes. After data acquisition, the images were pre-processed to remove the temperature scales and other marks and brought to the same dimensions by resizing and cropping. After this, various techniques were applied to augment the collected data to increase the number of images in the dataset. Finally, classification using convolutional neural networks (CNN) models was done and the results were compared. Fig. 2 depicts the workflow of the proposed work.

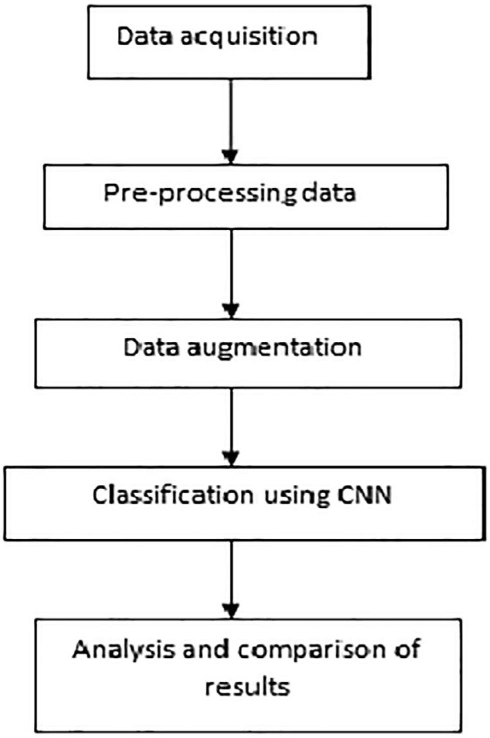



**Fig. 2.** Workflow of the proposed work.

### 4.1. Data acquisition

A thermographic camera (also called an infrared camera or thermal imaging camera) is a device that forms an image using infrared radiation, similar to a common camera that forms an image using visible light. Instead of the 400–700 nano-meter range of the visible light camera, infrared cameras operate in wavelengths as long as 14,000 nm (14 µm) (Thermographic_camera, 2018).

The higher an object's temperature, the more infrared radiation is emitted as black-body radiation. A special camera can detect this radiation in a way similar to the way an ordinary camera detects visible light. It works even in total darkness because the ambient light level does not matter. This makes it useful for rescue operations in smoke-filled buildings and underground areas (Thermographic_camera, 2018).

In the proposed work, we have used the FLIR ONE thermal camera for data acquisition. This camera displays live thermal infrared images using the FLIR ONE app. The advanced and patented multi-spectral dynamic (MSX) technology merges and extracts details from thermal images and visible images to create enhanced pictures and videos that reveal what the naked eye can't see.

Various places in Chandigarh city where potholes were available, were visited and pictures of roads with and without potholes were clicked. Along with the thermal images, the corresponding vision camera images of the pothole and non-pothole roads were also saved to get a clear view of the actual condition. Figs. 3 and 4 show sample images of the pothole and non-pothole roads captured using FLIR ONE and their corresponding vision images.

The following types of potholes were identified to collect images (Kim and Ryu, 2014):

- Based on the severity of pothole: Low, medium, and high
- Based on the presence of shade: Shady and non-shady
- Based on the presence of water: Water-filled, wet, and dry

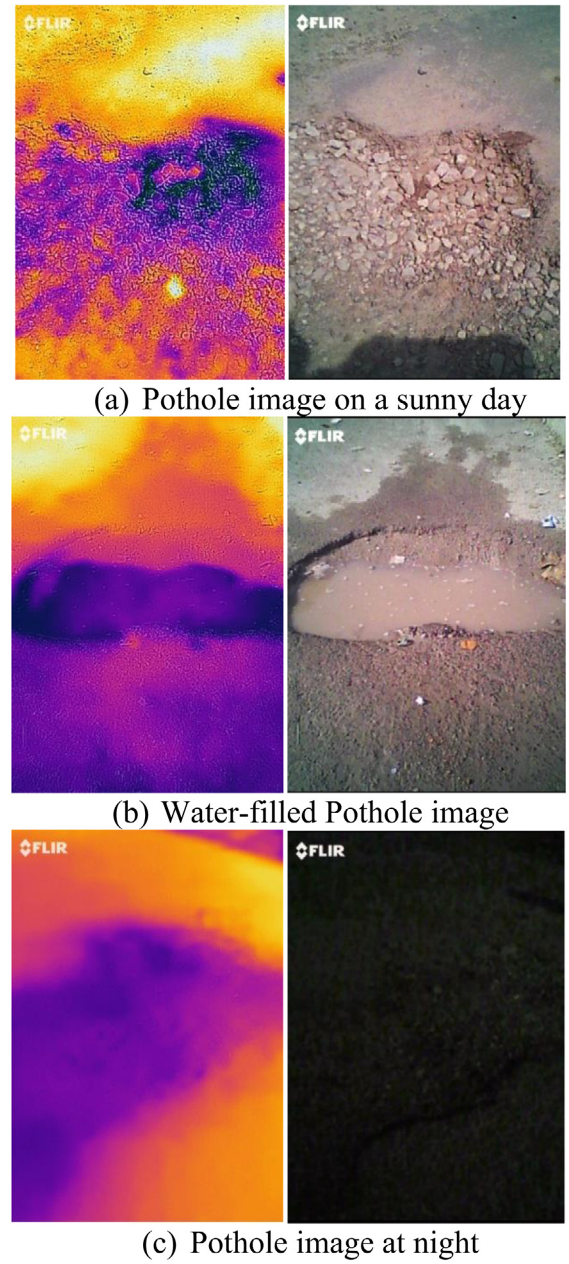



**Fig. 3.** Thermal and vision camera images of potholes.

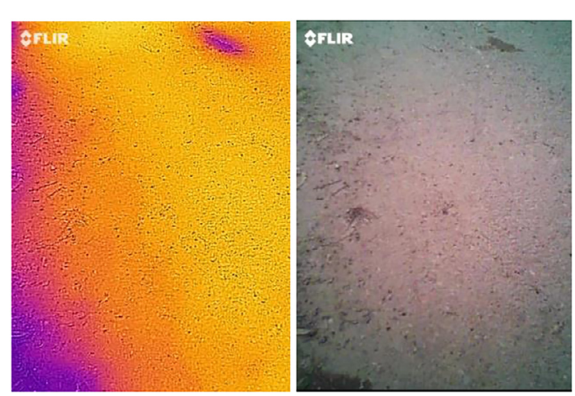



**Fig. 4.** Thermal and vision camera images of non-potholes.

Also, data were collected at different times of day and year to study the effect of light and temperature on the performance of thermal imaging in pothole detection.

*4.1.1. Parameters recorded*

While doing data collection, we prepared separate spreadsheets for recording the information of dataset from each of the two thermal cameras. Both spreadsheets had the following attributes:

- Thermal image id
- Vision image id
- Air temperature
- Road temperature
- Pothole temperature
- Time
- Severity
- Water filled
- Shade
- Location

Approximately 500 images of potholes and non-pothole roads were collected manually.

*4.2. Data pre-processing*

After the collection of images of potholes and non-potholes, we performed the pre-processing of the collected image dataset. During pre-processing of the dataset, the following various operations were performed on the images.

*4.2.1. Cropping of images*

Images were cropped to remove the default camera labels so that they do not hinder the image processing techniques. The dimensions of the thermal images captured after cropping were $480 \times 590$ pixels using FLIR ONE.

*4.2.2. Resizing of images*

After trying to run the CNN model on the cropped images, the hardware (GPU memory) was found not be enough to handle images of the above-mentioned pixels and 4500 images (after augmentation), therefore the images were resized to half i.e. $240 \times 295$ pixels before training.

*4.3. Data augmentation*

In a lot of real-world scenarios, even small-scale data collection can be extremely expensive or sometimes nearly impossible. Therefore, to make the most out of little data, various augmentation techniques have been used. This helps in better training, prevents overfitting and helps the model generalize better due to a larger dataset.

*4.3.1. Zooming*

Image zooming means converting the image into a magnified image. The zoomed version of thermal images can be used for augmentation as they represent the case when images are taken from a close distance from a pothole.

*4.3.2. Rotation*

Rotation of an image means changing the position of an object about a pivot point at some angle. Image of a pothole or non-pothole when rotated, will still look like a pothole or non-pothole and would represent the case as if the picture was taken from a different angle.

*4.3.3. Mirroring*

Mirroring creates a mirror image of the object. The mirror image of a pothole or non-pothole also represents a real scenario.

*4.3.4. Blurring*

The blurring of an image means decreasing the intensity of sharp features of the image like edges, corners etc. The thermal camera can also capture a blurred image on a foggy day. Thus, making the blurred image out of an original thermal image helps in increasing the dataset size and leads to better training of the model.

*4.3.5. Contrast enhancement*

Contrast enhancement helps in increasing the contrast in the image and thus making the image contain more intensities than just a few. This was applied on thermal images for augmentation since the same image can have better contrast on a sunny day or bright day.

*4.3.6. Salt and pepper noise*

Salt and pepper noise is added to an image by randomly changing intensities of some of the pixels of an image to 1 and some other to 0. Salt and pepper noise can represent a case where images were collected on a dusty day, or when the camera has dust on it.

*4.4. Classification of images using convolutional neural networks*

For applying CNN on the dataset, we had the option of developing our own model or using a pre-trained model. Thus, we performed various experiments by applying CNN in different ways to the dataset. The two broad experiments conducted were as follows:

1) Applying self-built CNN models
2) Applying CNN based ResNet models

ResNet (Residual Network) models were chosen as they are trained on ImageNet dataset, that contains a very large number of images of various objects. Also, these networks have performed fairly in various deep learning competitions with very less error percentage. They also solve the problem generally observed in large networks that is with increase in depth of networks, saturation and degradation in accuracy start taking place.

*4.4.1. Applying self-built CNN model*

In experiment 1, we tried to minimize the overfitting in training and to improve the accuracy rate. For this experiment, we also resized the data from dimensions 480x590 to 240x295, as it was becoming difficult to work with such a large size of images. We divided the dataset into train and validation set in the ratio of 90:10 after keeping 104 images separately for testing purpose. The parameters taken during the experiment are given below.

Train-validation split: 90:10
Image size: $240 \times 295$
Total categories: 2.
Total images: 4904
Training dataset size: 4320
Validation dataset size: 480
Test dataset size: 104
Kernel: $3 \times 3$ for convolution layers
Activation: ReLU for convolution layers
Loss function: binary_crosss entropy

*4.4.2. Applying CNN based ResNet models*

For experimentation, we split the dataset for training and validation in three different ratios of 60:40, 80:20 and 90:10. Then, we tried different ResNet models for experimentation such as ResNet18, ResNet34, ResNet50, ResNet101 and ResNet152. We then tested the models with these three dataset sizes by varying the image size. We experimented with four image sizes 224x224 (the standard ImageNet size), 240, 360, 480 (maximum size of the image).

### *4.5. Comparison of models*

After performing all the experiments, the following comparisons were done.

- Comparison between self-built and pre-trained models
- Comparison of different pre-trained models
- Comparison with previous existing work

The results obtained from these various comparisons are discussed in the next section.

## 5. Results and discussion

### *5.1. Applying self-built CNN model*

Experiment 1.1: Adding batch normalization layer after the last layer only.

Parameters:

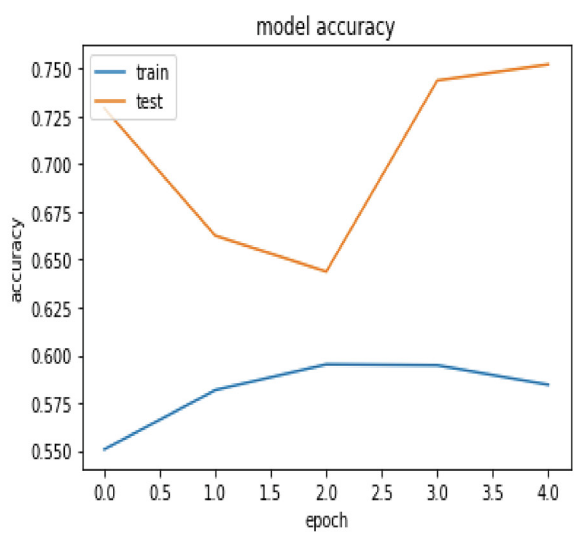
1) Batch normalization after the last layer
2) 4 convolution layers with output size 16,32,64,128
3) dropout 0.7 in the dense layer and sigmoid activation
4) learning rate = 0.001, decay = 0.0
5) No. of epochs = 5
6) Batch size = 32

Fig. 5 graphically represents the results of this experiment.
Observations

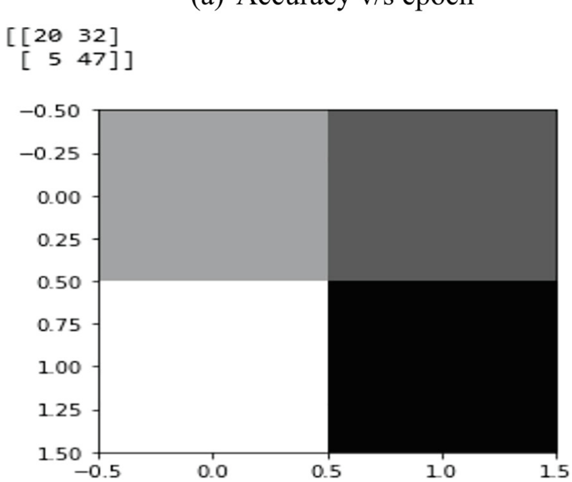
1) From this experiment, we achieved an average training accuracy of 58.12% and average validation accuracy of 70.62%.
2) The training and validation losses were on the higher side.
3) Test accuracy achieved was 64.42%

Experiment 1.2: Adding batch normalization layer after every layer.
Parameters changed:

1) Batch normalization after every layer
2) Learning decay = 0.01
3) Epochs = 10
4) Batch size = 16

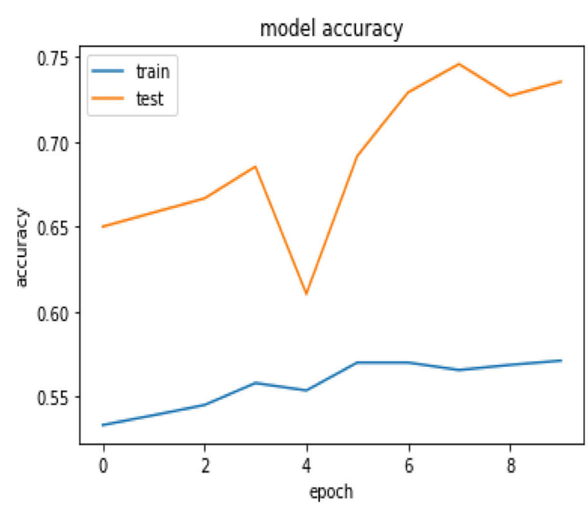



(a) Accuracy v/s epoch

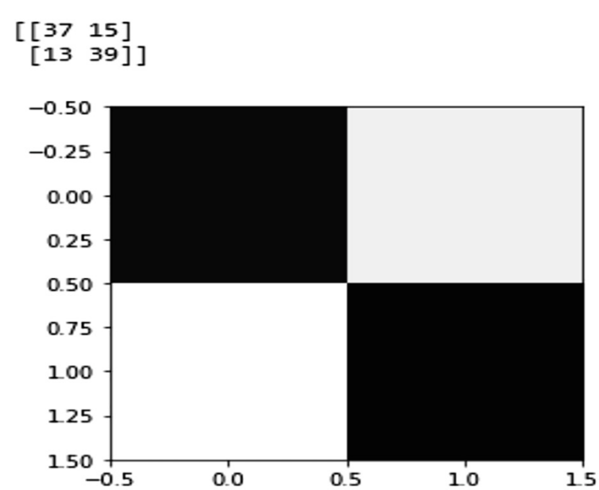



(b) Confusion matrix

**Fig. 5.** Results of experiment 1.1.




(a) Accuracy v/s epoch




(b) Confusion matrix

**Fig. 6.** Result of experiment 1.2.

**Table 2**
Best results obtained in each model in Experiment 2.1.

| S. No | Model | Image size | Training loss (%) | Validation loss (%) | Training accuracy (%) | Validation accuracy (%) | Average precision-recall score |
|---|---|---|---|---|---|---|---|
| 1 | ResNet18 | 224 | 30.18 | 27.37 | 88.69 | 90.52 | 0.88 |
| 2 | ResNet34 | 240 | 28.41 | 27.57 | 89.42 | 89.42 | 0.86 |
| 3 | ResNet50 | 224 | 25.11 | 24.07 | 90.05 | 91.77 | 0.89 |
| 4 | ResNet101 | 240 | 23.47 | 22.40 | 91.14 | 92.50 | 0.91 |
| 5 | ResNet152 | 224 | 28.16 | 23.28 | 90.20 | 91.66 | 0.89 |

Fig. 6 graphically represents the results of this experiment.
Observations

1) From this experiment, we achieved an average training accuracy of 55.74% and average validation accuracy of 68.99%.
2) The training and validation losses were on still higher side.
3) Test accuracy achieved was 73.06%

We saw that accuracy achieved was not good enough from these two experiments on self-built CNN and also, losses were high. Due to the unsatisfactory results in our model, we tried pre-trained ResNet models.

### *5.2. Applying CNN based ResNet models*

Experiment 2.1: Different models of ResNet with train-validation split 60:40 and different image sizes.

1. ResNet18

We tried fine-tuning this model on image sizes: 224*224, 240*240, 360*360 and 480*480. Best validation accuracy of 90.52% was achieved with image size of 224. Validation loss was 27.37%. Training the model on image size 480 led to out of memory error.

2. ResNet34

Best results were obtained from image size 240 with a validation accuracy of 89.42% and validation loss of 27.57%. An image size of 480 again led to a runtime error.

3. ResNet50

Best results were obtained from image size 224 with a validation accuracy of 91.77% and validation loss of 24.07%. An image size of 360 and 480 led to out of memory error in this case.

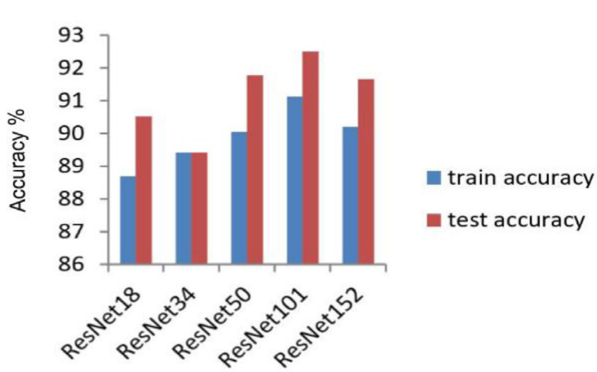



**Fig. 7.** Graph showing the best train and validation accuracy achieved by different models.

4. ResNet101

Best results were obtained from image size 240 with a validation accuracy of 92.50% and validation loss of 22.40%. An image size of 360 and 480 led to out of memory error in this case.

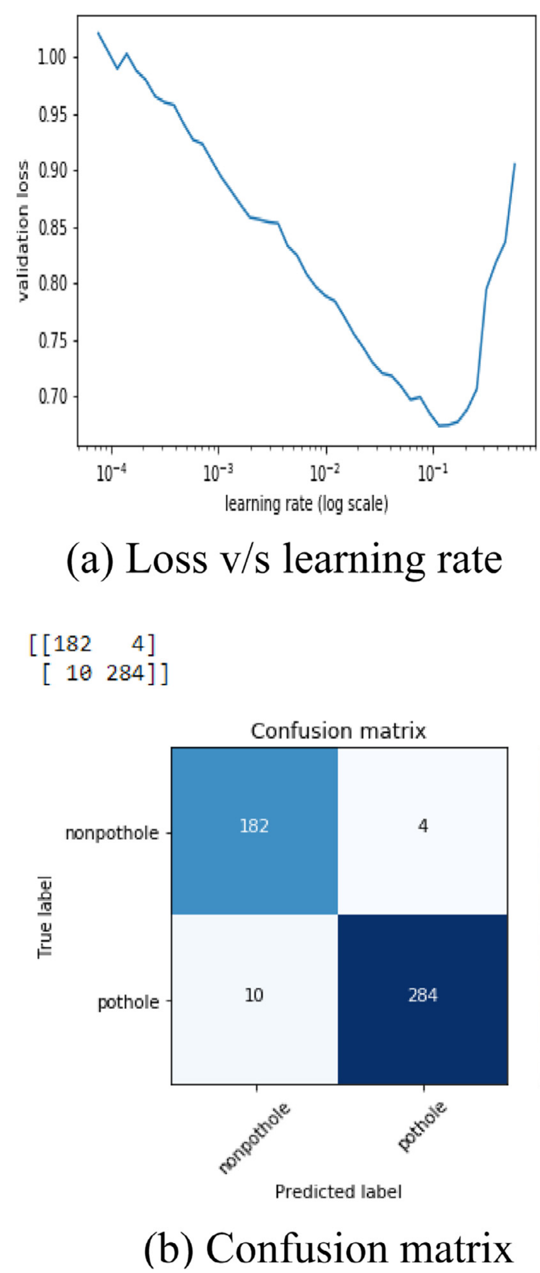



**Fig. 8.** Results of Experiment 2.3 with ResNet101 on 224x224 image size.

5. ResNet152

Best results were obtained from image size 224 with a validation accuracy of 91.66% and validation loss of 23.28%. An image size of 360 and 480 led to out of memory error in this case.

Table 2 summarizes results from all models used. From experiment 2.1, we can observe that average train accuracy of 89.9% and average validation accuracy of 91.17% is achieved with dataset divided in ratio 60:40 for train and validation.

For the further experiments, we worked on the top 3 performing models (shown in Table 2), that achieved an accuracy of 90% and above in training and validation, since with more data the accuracy increases proportionally. Also, we did not perform further experiments with image size 360 and 480, since they always gave out of memory error. Fig. 7 represents a comparison between train and test accuracies of different models.

Experiment 2.2: Different models of ResNet with train-validation split 80:20 and image size 224 and 240.

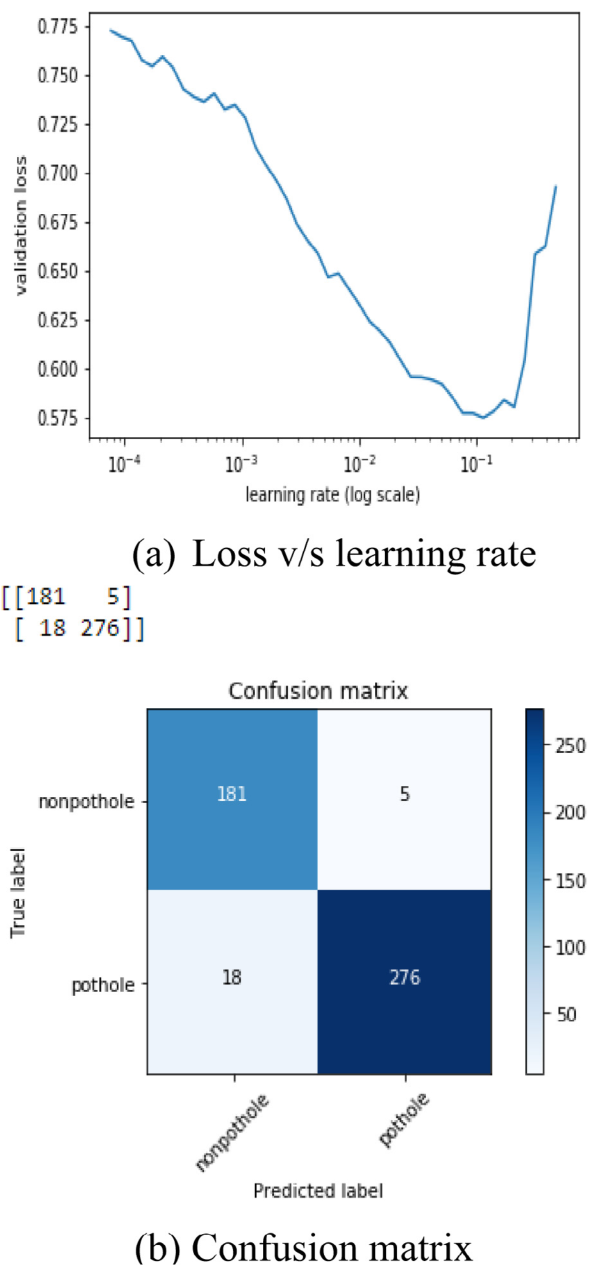



**Fig. 9.** Results of Experiment 2.3 with ResNet101 on 240x240 image size.

1. ResNet50

An image size of 240 resulted in a validation accuracy of 95.20% and validation loss of 20.05%. An image size of 224 resulted in a validation accuracy of 94.47% and validation loss of 13.13%.

2. ResNet101

An image size of 240 resulted in a validation accuracy of 93.95% and validation loss of 16.46%. An image size of 224 resulted in a validation accuracy of 94.16% and validation loss of 16.65%.

3. ResNet152

An image size of 240 resulted in a validation accuracy of 95.00% and validation loss of 14.89%. An image size of 224 resulted in a validation accuracy of 93.95% and validation loss of 14.06%

Observations

1) From experiment 2, we observe that accuracy increases with the increase in data.
2) The accuracies on image size 224 and 240 were almost similar
3) Average training accuracy achieved was 93.76 and average validation accuracy was 94.45

**Table 3**
Accuracy comparison between self-built and pre-trained models.

| Accuracy | Self-built model | ResNet |
|---|---|---|
| Avg. training | 62.63 | 94.64 |
| Avg. validation | 69.8 | 95.2 |

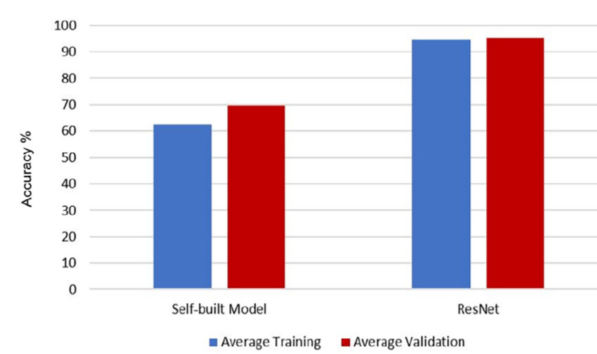



**Fig. 10.** Graph showing a comparison between self and pre-trained models in terms of accuracy.

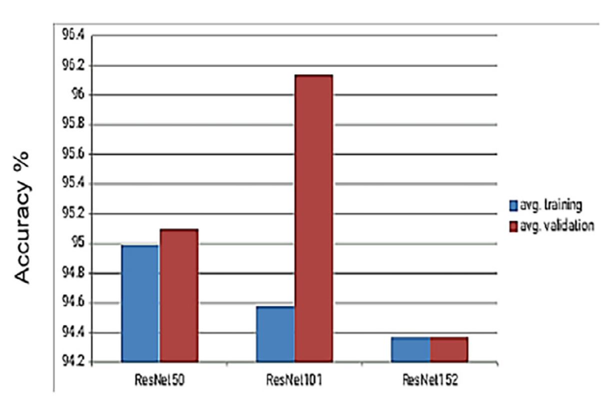



**Fig. 11.** Graph showing the comparison of different ResNet models.

**Table 4**
Comparison with previous works.

| S. No. | Reference | Method used | Dataset | Accuracy |
|---|---|---|---|---|
| 1 | Hoang et. al. (2018) | LSSVM and NN using steerable filter based feature extraction | Visual images of road samples | 87* |
| 2 | Ryu et. al. (2015) | Segmentation, candidate extraction, and decision | 2D road images collected from a camera | 73.5 |
| 3 | An et. al. (2018) | Various ANNs | Smartphone camera images of potholes | 96.75* |
| 4 | Proposed work | CNN based ResNet model | Road thermal images | 97.08 |

* - Average accuracy obtained from different models or methods.

Experiment 2.3: Different models of ResNet with train-validation split 90:10 and image size 224 and 240.

1. ResNet50

Best results were obtained from image size 224 with a validation accuracy of 95.62% and validation loss of 10.79%.

2. ResNet101

Best results were obtained from image size 224 with a validation accuracy of 97.08% and validation loss of 11.61%. This was by far the best accuracy achieved.

Fig. 8 and Fig. 9 graphically represent the results of this experiment.

3. ResNet152

Best results were obtained from image size 240 with a validation accuracy of 94.79% and validation loss of 16.28%.

Observations

1) The accuracies on image size 224 and 240 were almost similar.
2) A validation accuracy of 97.08% was observed, which was highest till now.
3) Average training accuracy achieved was 94.64 and average validation accuracy was 95.20.

*5.3. Comparison of results*

*5.3.1. Comparison between self-built and pre-trained models*

Table 3 shows the comparison between the results of the self-built model and the pre-trained models.

Fig. 10 shows the comparison between average accuracies of self and pre-trained models.

*5.3.2. Comparison among various pre-trained models*

Since the best results were obtained in case of Experiment 2.3 discussed above with training and validation ratio of 90:10, thus, we compare the different models on the basis of results obtained on this dataset. Fig. 11 shows the comparison among average accuracies obtained by different pre-trained models.

*5.3.3. Major findings from the experiments*

The above experiments reveal that CNN based ResNet models clearly outperform the self-built CNN models. An average accuracy of 95% has been achieved using ResNet models. ResNet50 performed consistently well in each case but ResNet101 outperformed ResNet50 in one case with 97.08% accuracy. This means that both ResNet50 and ResNet101 can be a suitable choice of models for this problem.

Also, best image dimension to work with is 224x224 while using ResNet models, as it gave good results in each case. Use of cyclic learning rates has improved accuracy a lot. Best accuracy achieved with ResNet101 is also better than many of the previous accuracies obtained till now. Also, we can analyze from the confusion matrix that the false positive rate is low.

*5.3.4. Comparison with previous existing work*

Since the use of thermal imaging in the field of pothole detection is relatively a new concept. Not much work exists in this direction which can be compared. Also, results of the previous works using different methods other than thermal imaging can be used for comparison, but it should be kept in mind that these works used their own datasets and algorithms for pothole detection. Thus, the only common ground on which we can do a comparison is that they all are doing pothole detection. Table 4 compares the proposed technique with some previous works.

## 6. Conclusion

The pothole detection using artificial intelligence methods can help in better maintenance of the road conditions especially in developing countries where resources are limited. For this purpose, the proposed system based on convolutional neural networks using thermal imaging does have the potential to compete with the existing techniques of pothole detection. The proposed pothole detection system using CNN based ResNet model has achieved an accuracy of 97.08 which is the best ever reported in the literature so far. Moreover, this is the first time when thermal imaging has been used for the pothole detection which has several advantages over the other techniques such as more accurate, low cost, less complex, can work in night and foggy weather conditions and also does not involve risk of passing through potholes.

Further, this work can be extended to detect the region of potholes after classifying an image as pothole and furthermore parameters can also be detected like the severity of potholes on the basis of which it can be figured out which area requires urgent repair work.

## Acknowledgment

This work is carried out in collaboration with the Design Innovation Center, Panjab University, Chandigarh, India.

## References

Kim, T., Ryu, S.K., 2014. Review and analysis of pothole detection methods. J. Emerging Trends Comput. Information Sci. 5 (8), 603–608.
Kim, T., Ryu, S.K., 2014. A Guideline for Pothole Classification. Int. J. Eng. Technol. 4 (10)
Thermographic_camera [Online] https://en.wikipedia.org/wiki/Thermographic_camera (accessed on 15th November, 2018)
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
Hoang, N.D., 2018. An Artificial Intelligence Method for Asphalt Pavement Pothole Detection Using Least Squares Support Vector Machine and Neural Network with Steerable Filter-Based Feature Extraction. Adv. Civil Eng.
An, K.E., Lee, S.W., Ryu, S.K., Seo, D., 2018, January. Detecting a pothole using deep convolutional neural network models for an adaptive shock observing in a vehicle driving. In: Consumer Electronics (ICCE), 2018 IEEE International Conference on (pp. 1-2).
Ryu, S.K., Kim, T., Kim, Y.R., 2015. Image-based pothole detection system for ITS service and road management system. Math. Problems Eng. 1, 2.

Negied, N.K., 2014. Human Biometrics: Moving Towards Thermal Imaging. Int. J. Recent Technol. Eng. (IJRTE), ISSN, 2277–3878.

Dash, D.K., 2018. Potholes killed 3,597 across India in 2017, terror 803 [Online] https://timesofindia.indiatimes.com/india/potholes-killed-3597-across-india-in-2017-terror-803/articleshow/64992956.cms (accessed on 20th November 2018)

Akagic, A., Buza, E., Omanovic, S., 2017, May. Pothole detection: an efficient vision-based method using RGB color space image segmentation. In: Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2017 40th International Convention on (pp. 1104-1109). IEEE.

Yu, B.X., Yu, X., 2006. Vibration-based system for pavement condition evaluation. In: Proceedings of the 9th International Conference on Applications of Advanced Technology in Transportation, pp. 183–189.

De Zoysa, K., Keppitiyagama, C., Seneviratne, G.P., Shihan, W.W.A.T., 2007. A public transport system based sensor network for road surface condition monitoring. Proceedings of Workshop on Networked Systems for Developing Regions, 1–6.

Erikson, J., Girod, L., Hull, B., 2008. The pothole patrol: using a mobile sensor network for road surface monitoring. In: Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services (2008), pp. 29-39

Mednis, A., Strazdins, G., Zviedris, R., Kanonirs, G., Selavo, L., 2001. Real-time pothole detection using Android smartphones with accelerometers. In: Proceedings of the International Conference on Distributed Computing in Sensor Systems and Workshops, pp. 1–6.

Wang, K.C.P., 2004. Challenges and feasibility for comprehensive automated survey of pavement conditions. In: Proceedings of 8th International Conference on Applications of Advanced Technologies in Transportation Engineering, pp. 531–536.

Chang, K.T., Chang, J.R., Liu, J.K., 2005. Detection of pavement distress using 3D laser scanning technology. In: Proceedings of the ASCE International Conference on Computing in Civil Engineering, pp. 1–11.

Hou, Z., Wang, K.C.P., Gong, W., 2007. Experimentation of 3D pavement imaging through stereovision. In: Proceedings of the International Conference on Transportation Engineering, pp. 376–381.

Li, Q., Yao, M., Yao, X., Xu, B., 2009. A real-time 3D scanning system for pavement distortion inspection. Measurement Sci. Technol. 21 (1), 15702–15709.

Staniek, M., 2013. Stereo vision techniques in the road pavement evaluation. In: Proceedings of the XXVIII International Baltic Road Conference, pp. 1–5.

Joubert, D., Tyatyantsi, A., Mphahlehle, J., Manchidi, V., 2011. Pothole tagging system. In: Proceedings of the 4th Robotics and Mechatronics Conference of South Africa (2011), 1-4.

Moazzam, I., Kamal, K., Mathavan, S., Usman, S., Rahman, M., 2013. Metrology and visualization of potholes using the Microsoft Kinect sensor. In: Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (2013), 1284-1291.

Koch, C., Brilakis, I., 2011. Pothole detection in asphalt pavement images. Adv. Eng. Inf. 25, 507–515.

Jog, G.M., Koch, C., Golparvar-Fard, M., Brilakis, I., 2012. Pothole properties measurement through visual 2D recognition and 3D reconstruction. In: Proceedings of the ASCE International Conference on Computing in Civil Engineering, pp. 553–560.

Lokeshwor, H., Das, L.K., Sud, S.K., 2013. Method for automated assessment of potholes, cracks, and patches from road surface video clips. Procedia – Social Behav. Sci. 104, 312–321.

Koch, C., Jog, G.M., Brilakis, I., 2013. Pothole detection with image processing and spectral clustering. J. Comput. Civil Eng. 27 (4), 370–378.

Buza, E., Omanovic, S., Huseinnovic, A., 2013. Stereo vision techniques in the road pavement evaluation. In: Proceedings of the 2nd International Conference on Information Technology and Computer Networks, pp. 48–53.

Lokeshwor, H., Das, L.K., Goel, S., 2014. Robust method for automated segmentation of frames with/without distress from road surface video clips. J. Transp. Eng. 140, 31–41.

Yamashita, R., Nishio, M., Do, R., Togashi, K., 2018. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9(4), 611–629. Advance online publication. doi:10.1007/s13244-018-0639-9

Rajaraman, S., Antani, S.K., Poostchi, M., Silamut, K., Hossain, M.A., Maude, R.J., Thoma, G.R., 2018. Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. PeerJ 6, e4568.

Roberts, C., V., 2018. Quantifying the extent to which popular pre-trained convolutional neural networks implicitly learn high-level protected attributes, A master's thesis, Department of Computer Science. Princeton University.

Yang, C.-Y., Zhang, Y.-D., Xiao-Jun, Y., 2019. Exact solutions for the differential equations in fractal heat transfer. Thermal Science 20 (Suppl. 3), S749–S752.

Wang, S.H., Muhammad, K., Phillips, P., et al., 2017. Ductal carcinoma in situ detection in breast thermography by extreme learning machine and combination of statistical measure and fractal dimension. J. Ambient Intell. Human Comput. https://doi.org/10.1007/s12652-017-0639-5.