Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 10, 2025

CRACKS DETECTION IN WALLS USING MACHINE LEARNING TECHNIQUES

Muhammad Jawad Khan[1], Sahib Khan[2], Syed Waqar Shah[3], Bilal Ur Rehman[*4], Muhammad Amir[5], Humayun Shahid[6], Kifayat Ullah[7]

[1,2]*University of Engineering and Technology, Mardan, Pakistan*
[3,*4,5,7]*Department of Electrical Engineering, University of Engineering and Technology, Peshawar, Pakistan*
[6]*Department of Telecommunication Engineering, University of Engineering and Technology, Taxila, Pakistan*

DOI: https://doi.org/10.5281/zenodo.17627828

Keywords
Machine learning, Deep Learning, CNN, Cracks detection, No-cracks detection

Article History
Received: 11 September 2025
Accepted: 21 October 2025
Published: 04 November 2025

Corresponding Author: *
Bilal Ur Rehman

Abstract
This paper presents a method for detecting cracks in walls using machine learning and deep learning techniques. Here are two types of image datasets that contain cracked images and non-cracked images. The dataset is named 'positive' and 'negative'. The positive dataset contains images with some cracks, and the negative dataset has images without cracks. Machine learning techniques are used for detecting cracks in the images. Machine learning encompasses three primary learning techniques: unsupervised learning, reinforcement learning, and supervised learning. In supervised learning, the user trains a program on known and labeled data, while in unsupervised learning, the user trains a program on unknown and unlabeled data. In this paper, we have examined unknown and unlabeled data of wall images. We have trained a program using a convolutional neural network to identify patterns within the data. A CNN is used to learn patterns within images, determining whether an image contains a crack or not. A CNN is a class of artificial network that plays an important role in various computer vision tasks. CNN uses the backpropagation concept by applying multiple building blocks, such as convolutional layers, Rectified Linear Units (ReLU), pooling layers, and fully connected layers.

INTRODUCTION

The preservation of old-aged buildings and the restoration of damaged buildings are crucial for their survival and prolongation of their remaining lifespan. The damage in the building occurs when the following factors become apparent, such as typical damage due to material losses, changes in weather, algae growth, and dust deposition. The sudden and quick earthquake causes cracks in buildings, houses, traditional buildings, and historic places. The cracks in the houses appear very different from those in the parts of the building where the preservation of old buildings is meaningful for economic development through tourism. Some historic buildings become increasingly vulnerable to collapse due to earthquakes, heavy rain, fire, seepage of water, and weather changes, which cause degradation of the building materials. Environmental factors such as these can also lead to cracks in the building.

On the other hand, preservation of cracks is a prime part of their remaining service life. The most significant factor contributing to cracks is moderate and occasional earthquakes, which cause small and large cracks in buildings. The monitoring and preservation of cracks in buildings also play a crucial role in their survival.

We have been analyzing and observing the results of experiments on both cracked and non-cracked datasets. The model we selected for detecting cracks in wall images is the convolutional neural network. The machine learning techniques applied for training the model to abstract the information and features inside the images to detect and predict whether the following image has some cracks and the rest of the image has no cracks based on the trained model [1]. Now, we introduce the concepts of machine learning and deep learning, along with their differences, and the applied model of deep learning in this paper. How is the machine learning model and how is the machine learning model on the images of walls that consist of two parts, such as the images of walls and images of walls have no cracks. The machine learning model is trained on 40,000 images, which is a combination of cracked and non-cracked images of walls. Machine learning refers to any computer program that "learn" patterns and predicts by itself without having to be explicitly programmed by a human. The goal of machine learning is to make machines act more like humans, because the more intelligent they become, the more they can help us. We use computers to process large amounts of data for machine learning, where computers learn from the data and make predictions and decisions. The primary aim of machine learning is to make predictions about data, such as predicting the type of thing, classifying objects within an image, or making predictions on multi-class classification problems.

All these tasks as the same as what humans predict by thinking and learning about something: data scientists are expected to be familiar with the differences between supervised machine learning and unsupervised machine learning. Our problem is detecting cracks in images, and the model is trained in data that is not in a structured or ordered form. The model is trained on unsupervised data because, in unsupervised machine learning, the algorithms generate an answer based on the unknown and unlabeled data. Our data is related to unknown and unlabeled data. The supervised learning-related problem involves categorized data and its algorithms, such as classification and regression, including random forests and decision trees. All these types of algorithms are used for supervised learning. As such, the supervised learning prediction is based on the structured data. The Unsupervised learning problem has not categorized data as such as supervised learning has. There must be some algorithms that will be used for the problem, like Unsupervised learning in machine learning, such as Association rule learning, Clustering, K-means, etc. The unsupervised data includes images, audio recognition, face recognition, and object recognition in images. The Reinforcement of data is related to skill acquisition and real-time learning. The Reinforcement of data based on positive and negative feedback leads to improvements in the model's performance and degradation of its performance. Such models are used in automated games [2].

Literature review

A significant number of trails were previously carried out by researchers aimed at recognition as well as identification of defects in object and cracks in buildings and other civil structures, with particular interest focused on buildings with historical heritage and similar archaeological value. Different techniques and approaches are applied for feature extraction and identifying discontinuities within the images for further experimentation. Hoang formulated a technique that focused on localization of faults and defects within the digital images of structural walls by deploying techniques drawing heavily from both machine learning and image processing, mentioned earlier on in the introduction section. The machine algorithms are illustrated according to problems related to machine learning, including supervised learning classification, regression, random forest, and decision trees, as well as unsupervised learning problems such as association rule learning and clustering, including k-means. Hoang, in it, the features had been extracted by applying the steering filters and the projection filters. In different countries, surveys of high-rise buildings are usually performed based on visual inspections by human inspectors. In this regard, the potential problems in the buildings are directly pointed out by visual bases, which makes it difficult for qualified technicians to make changes to the buildings' potential structures and make informed decisions about the cracked buildings. There must be some disadvantages in the human visual assessment surveys of high-rise buildings. First, the surveying process is significantly affected by the subjective judgment of inexperienced individuals. Second, in the human visual assessment process, measuring, processing, and reporting data about the cracks found in high-rise buildings must be time-consuming to cover all sides for inspection [3].

A method was presented that focused on classification followed by immediate feature extraction corresponding to cracks that manifest along various pavements by Chen et al. The team proceeded by applying a mean filtration technique for doing away with noise all while not damaging the features that are distinctive and prime for the same set of images. For feature extraction, the technique adhered to was local binary pattern, followed by deployment of support vector machines aimed at categorization of the pre-segmented images into categories such as crack or no crack, and further into sub-categories such as a transverse crack, alligator or longitudinal crack, and so forth. The team concluded that the aforementioned combination of techniques lead to securing of classification and detection accuracy when it comes to marking alligator cracks. Next in line are the transverse cracks. Finally, longitudinal cracks land at the end in terms of attainable classification efficacy [4].

Likewise, Quan et al. also came up with an improved technique aimed at detecting cracks present in images presented an improved method for crack detection based out of the fine details manifesting across grey-level histograms of relevant images. The proposed technique makes use of non-linear media filtering to remove residual noise and for accommodation of oil stains and other obscuring spots on the road and pavements. With this technique, a high level of accuracy has been achieved and reported in the relevant work [5]. Nedunuri came up with a interface based on digital graphics for diving deep into the cracking properties prevalent in concrete walls. presented a graphical user interface for the investigation of cracking properties in concrete walls. The model output worked impressively well for results along the length, width, and classic types of cracks. Velumani et al. proposed a method that finds cracks in buildings using an image processing-based approach. In 2019, a fully convolutional encoder and decoder neural network was presented for the semantic segmentation of concrete cracks. The model train and VGG16-based classifier were trained end-to-end for the recognition of cracking and non-cracking images [6]. Kim and Cho presented a method for crack detection and developed an assessment framework for concrete structures using a mask and a region-based convolutional neural network. In this model, Kim and Cho used the crack detection model for detecting and recognizing cracks, achieving a recall of 76.15%. They achieved better results by applying a crack detection-based model and a convolutional neural network [7].

The automated digital image-based model was proposed for finding cracks in the Road or pavement. The images were recognized and segmented in this model to find cracks in the pavement. A number of convolutional neural networks, eight to be specific, have been put into use for experimentation that lead up to detection and classification of cracks and their different types as they manifest across various pavements. The pre-trained convolutional neural networks are GoogleNet, SqueezeNet, ResNet-18, ResNet-50, ResNet-101, DenseNet-201, and Inception-v3. Based on sensitivity, accuracy, precision, and other factors, their performance is measured. SqueezeNet and GoogleNet achieved the highest classification performance as compared to the other pre-trained convolutional neural networks [8].

Cracks are also a significant problem for Engineers as well as building owners to identify whether some cracks occur internally in the buildings or externally within the concrete. The different hypotheses and observations have proven that as cracks are created and propagate, this also occurs over time. The occurring of cracks in the building gives us an essential sign of structural breakdown. The cracks occur due to the unusual earthquakes, low concrete use at the time of construction, less use of cement, poor knowledge, and inexperience at the time of building construction. In the detection of cracks, the Otsu process is a common technique for finding cracks in pavements, Roads, Buildings, and walls. However, there are some unsatisfactory aspects of the Otsu technique, including poor contrast, irregular shape, and a concrete appearance in the images. To overcome these disadvantages, the Min-Max Gray Level Discrimination (M2GLD) Image algorithm is proposed to improve the Otsu binarization procedure to increase the crack detection performance [9].

The defects in the railway wheel are distinguished as a significant origin in the foundation of the railway. The method is applied to recognize defects in the wheel, which is dependent on a technique that learns various types of defects and utilizes machine learning algorithms to anticipate whether a wheel has an imperfection or not during the activity. They are taking a lot more to take care of the system through visual bases and make their framework and assessment, but the accidents continue to happen again and again. The innovation played an important role in providing more accurate and procedural approaches for human beings. The most recent innovation has reached a point where, for example, Machine learning has been applied and employed in many fields, including engineering systems that interact with us and our daily lives. In this model, wheel defects are detected, which cause noise and vibration emission in the surrounding area. To overcome these defects in the railway wheel using machine learning algorithms, it can detect and recognize defects and solve them in the railway wheel. Any obstacle that is found on the path can be detected, and then an alert message is sent to the rail station to take immediate action as soon as possible. [10]

In this paper, seven machine learning algorithms are used and utilized for palmprint recognition. First, propose a technique or methodology for Region of Interest (ROI) feature extraction, which consists of two key techniques. In it, also applied 7 different supervised machine learning techniques were applied to recognize the palmprint. All the techniques have been tested, including Neural Network, Support Vector Machine, Native Bias, K-Nearest Neighbors (KNN), Random Forest, Decision Tree, and Adaptive Boosting. All techniques have been tested and utilized, and as a result, the MNs and KNN have been given the best recognition and accuracy for the palmprint of the region of interest [11].

Machine learning has a set of rules and a well-developed method for automatically gaining features and recognition capabilities applied to various fields and problems, including biomedical image analysis. In his research work, a plethora of machine learning techniques that are supervised in their nature are employed for the purpose of bringing about much needed improvements in cell segmentation, and also for the purpose of tracking attained via the active learning application. Primarily being a set of computational tools that are more dominant today than ever, machine learning largely remains one of the most beneficial developments in the field of math and computing taken together. Machine learning has been applied to various applications, including the detection of targets in images, the identification of people in images, route planning for car navigation systems, and many more. Machine learning algorithms are applied in this article to extract features and discover patterns, such as identifying clusters and phenotypes from cell appearances. There are two distinct flavors of machine learning applications. First, machine learning has been fully-embraced wherein a number of robotic agents working under the supervised learning mode of operation, have been actively deployed to help with processing of cells particularly in scenarios where automatic identification is sought across thousand of images. Second, Machine learning is used for unsupervised learning or as an analytical tool, such as visualization, clustering, and mathematical modeling. Machine Learning was originally put forward as technique for as a core technique for full-on implementation of various applications that shape up the evolving realm of intelligence of the artificial sort. In today's world, the main motivation behind developing artificial intelligence was to develop, create and program a robot that works like a human and thinks like a human brain, which laid the foundation for the jaw-dropping applications that we say today. Machine learning methods are divided into two categories: the first is supervised learning and the other is unsupervised learning. In supervised learning, the objects have correct labels, and a subset of the data set with labels is called the training set. In this method, the program and dataset are trained, and then they are used to predict new object labels for which the labels are not provided. It remains imperative to understand that as an outcome of supervised learning process, there generally are predesignated labels that are assigned to objects or artifacts that are similar to the type of entities appearing on the dataset. If the label takes only one value in a predefined trained value, this type of prediction is known as classification. If the label takes some continuous variables, then it is known as a regression. In machine learning, the most famous method and technique is linear regression. Taking into consideration the provided supervised model, and focusing on the training phase itself, the data used by the model for training is used as the basis for optimizing the parameters responsible for learning, thereby bringing about the much needed optimization that would, in turn, lead to classification of the data under investigation as close as possible to the one carried out during the training phase. In it, for every new data point, it asks for a set of rules and conditions that the particular feature exceeds a threshold or is below the threshold, and then the decision tree decides to select the option for yes or no. There is another method in supervised learning called the artificial neural network (ANN) that provides a prediction for new objects or instances based on predefined models, which also determines its flow of computation. The second flavor of machine learning is unsupervised learning. In this method, the common aim is to make predictions on the data, but this is the reverse of supervised learning, which deals with unlabeled data. Thus, a vast variety of tasks fall into this method, which are shown in the above statement. [12]

The most common form of machine learning is supervised learning. We want to build a system and then classify the images that have lots of objects to be detected, such as a person, a car, a dog, a road, an apple, and a motorbike. First, to collect a large data set of images of a car, dog, Road, apples, and a motorcycle, each labelled with its categories. After training, the machines produce an output in the form of a vector of scores. We want the desired category to achieve the highest vector score of all categories. A gap is created between the desired value and the output scores. To reduce this error, the machine applies its internal parameters. These internal or adjustable parameters, often called weights. Deep learning has a computational model composed of multiple processing layers that allow objects to pass through these layers, and as a result, important features are abstracted. Machine learning systems are used to identify objects in images, translate speech into text, and match new items, among other tasks. These applications utilize a class of techniques known as deep learning. The concept of a convolutional neural network (ConvNet) originated in discussions about deep learning. A convolutional neural network is designed to process data in the form of multiple arrays, such as a color image consisting of pixel intensities in three color channels. The image itself is in the form of three 2D arrays. There are multiple types of data, which are in the form of 1D, 2D, and 3D. The 1D data includes signals, sequences, and language. In 2D, there are images or audio, while in 3D, there are videos or volumetric images. There are four main ideas and steps involved in the convolutional neural network, which are: convolutional layer, Pooling layer, Rectified Linear Units (ReLU), and fully connected layers. In a convolution layer, one image becomes a stack of filtered images. The image is convolved with a bunch of filters or a bunch of features and creates a stack of filter images called a convolution layer. The result of the convolution layer and its weighted sum is then passed through a layer called the rectified linear unit.

In this layer, a stack or a bunch of images becomes a stack or a bunch of images with no negative values. The second layer is the pooling layer. In this layer, a stack of images becomes a stack of smaller images. First, choose the window size, usually 2 or 3. Pick a stride, usually 2. Then, walk a window across a filtered image and take the maximum value from each window. In the early 1990s, convolutional neural networks were experimented with for object recognition in images, including faces, hands, and face recognition. Until 2012, there had been numerous successes with ConvNets when a deep convolutional network was applied to millions of images taken from the web, which contained 1,000 different classes. After the spectacular results achieved from the deep convolutional network, this became the new era of deep ConvNets.

The success came from the use of GPUs and ReLUs. In the last layer, named the fully connected layer, there is a list of feature values that becomes a list of votes. Votes depend on how strongly a value can predict either this thing or the other. We fed them a picture x or y, and then it decides how strongly they predict the correct option and its probability. In a fully connected layer, the object is correctly classified by announcing its prediction using the method called backpropagation. Finally, all the statements show that convolutional neural networks only capture local "spatial" patterns in data if the data cannot be made to look like an image. ConvNets are less valuable. Convolutional Neural Networks (ConvNets) are excellent at identifying patterns and utilizing them to classify images [13].

The most developed algorithm among various deep learning models is the convolutional neural network (CNN), which is a class of artificial neural networks (ANNs) that are dominant methods in computer vision tasks. Some more deep learning networks, such as recurrent neural networks (RNN) for sequence models. Some terms are used during the training and application of convolutional neural networks, "parameters," which are variables that are learned during the training process. The term "hyperparameters" also refers to a variable that needs to be set before training a model. The "kernel" is the learnable parameters used during the convolution layer. As we already discussed, the different parts or layers of a convolutional neural network are mentioned in the above statements. There are mainly four building blocks in the CNN, which are the convolution layer, the rectified linear units (ReLUs) layer, the pooling layer, and the fully connected layer. The first three, convolution, rectified linear units ReLUs and pooling layers are used to extract the features from the input data or images.

What appears to be a stark contrast when stacked against what precedes, the third unit, wherein a fully-connected layer is set up, serves the purpose of declaring the extracted features as the definitive final output when considering the classification function. The first layer in the CNN, a convolutional layer, is a key layer that plays a crucial role in feature extraction. This layer is composed of a mathematical operation, such as a convolution (*) of the input image pixel and the kernel. The kernel is an optimizable feature extractor, which is applied to the input image at each position, making the CNN more efficient and effective for image processing [14]. A complete labelled diagram of a convolutional neural network is shown in Figure 1.

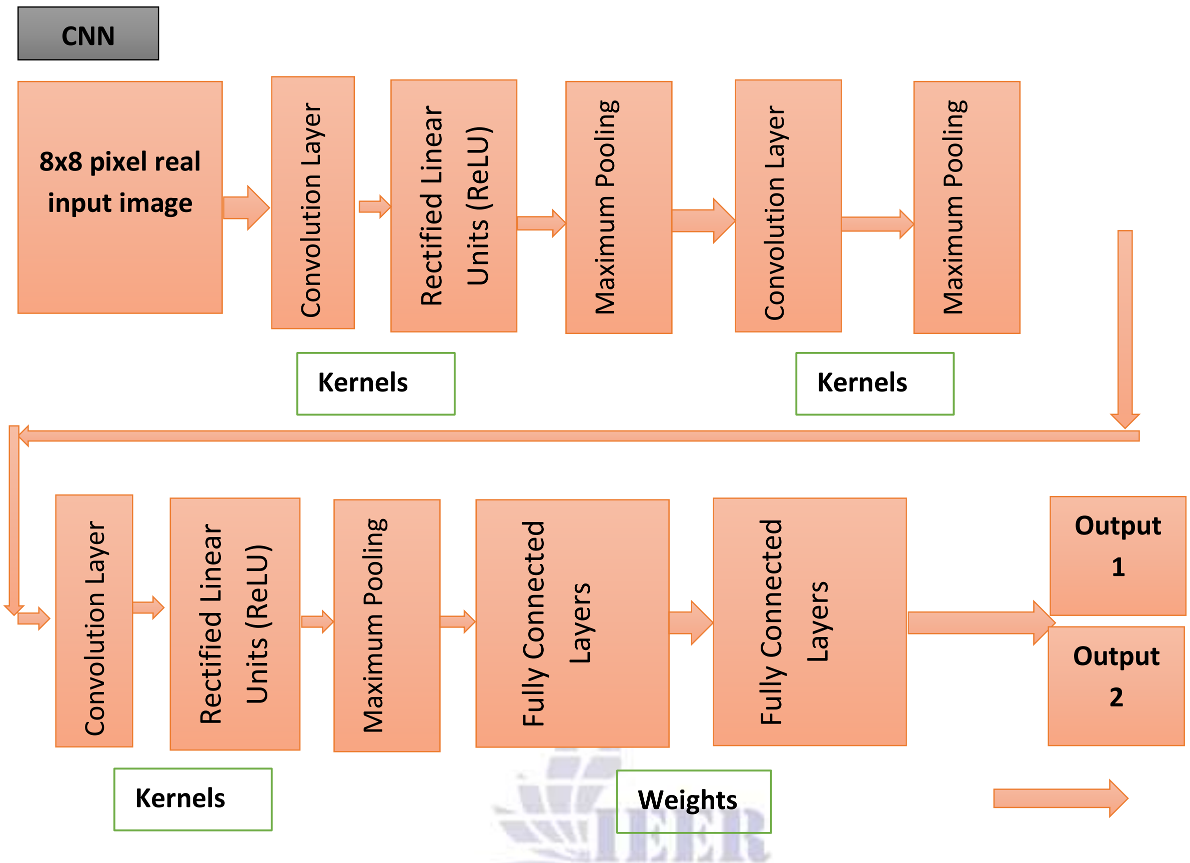


Figure 1: Overview of CNN Architecture and Its Training Process

Methodology:

Developed procedure

The main objective of the present study is to develop a method which is designed for the classification of cracks in the same infrastructure such as walls. The output of this method is the method, in which, it will be decided that whether or not the captured images contains cracks or no cracks or what type of cracks are present in the images. The overall model is trained and designed for recognition of cracks in the images of walls. The images of walls are downloaded from the Kaggle as form of dataset. The dataset contains two types of more datasets, which name as the positive and negative dataset. The negative dataset contains twenty thousand images of walls having no cracks, all are free of cracks. The positive dataset also contains twenty thousand of images having cracks in all of images. In this method, no pre-trained model is used from the web such like deep ConvNets. In this method and model, we used the google colab for our model to get free access of a graphical processing unit (GPU), from the web server. Google colab is just another type of jupyter notebook. The colab is a jupyter notebook running on one of the google servers which access a free GPU. We need a GPU in mode because when we are working with the unstructured data often times patterns are lots of vast and running a neural network a special kind of machine learning model, it does lot more computing power. The GPU is speed up the numerical computation and save a time. The GPU is a lot of fast and is doing the numerical analysis. The GPU is helps us to increase the speed of finding the patterns in the images. First, we made a folder name as surface crack detection in my drive to upload my datasets in the folder. Uploaded our dataset in the google drive to access all the data form google drive to the google colab. The data are uploaded successfully in the google drive with the extension (.zip), then we unzip all the images of negative and positive dataset. After that, all the important libraries of python used for machine learning problems are imported in the coding place such as numpy, pandas, matplotlib, seaborn, plotly.express, tensorflow, os, cv2, keras. Also, we import some modules like image from keras.preprocessing, train_test_split from sklearn.model_selection, the imported layers of CNN such as dense, conv2d, maxpool2d, flatten and dropout from the keras.layers, image Data Generator form the keras.preprocessing.image, the confusion_matrix, classification_report and r2_score from the sklearn.metrics etc. we made two directory a positive_dir and negative_dir for providing the path to the cracks images and no cracks images. we used the glob function for finding certain patterns in the data like images having the jpg extension. Also, we print the images by finding the correct number images in each dataset. The number of cracks images are found, and print form the dataset are twenty thousand in a number same as the case of no-cracks images. The python libraries, python libraries for machine learning and python libraries for deep learning are also shown in the order below:

Python Libraries:

- Numpy
- Pandas
- Matplotlib
- OS Module
- Plotly
- Seaborn
- CV2

Python library pandas used for making a data frames such as 1-dimensional and 2-dimensional data frame. The data frame consisting the information about object, images etc. The Numpy is used for making array of information. An array also consists of 1D and 2D array. A matplotlib is used for graphs to draw a given information and detail in a graph. The OS module in python provides the functions interacting with the operating system. The Plotly is an open-source python graphing library. Seaborn is used python data visualization based on matplotlib. CV2 is an open-source computer vision python library used for solving the vision recognition based problem.

Python Libraries for Machine Learning (Supervised Learning)

- Scikit-Learn

The python library Scikit-learn used for machine learning is used for predictive analysis. It is a massive library, there are tools, functionality and the ability that Sklearn shows.

Python Libraries for Deep Learning (Unsupervised Learning)

- Tensor flow
- Py-Torch
- Keras

The tensorflow is a numerical computing and deep learning library. Tensorflow is used to build deep learning and neural network model to gain and learn patterns inside an unstructured data. The tensorflow in this model is used for building a deep learning deep learning model and applied the convolutional neural network for extracting the features in the images of cracked and non-cracked images of walls. These libraries are used in this model for different tasks and for various computational analysis. Now the time to explain each library for itself working in the crack detection in images by using machine learning techniques and methodology.

Method Implementation

The developed method in this paper is applied by using the dataset of surface crack detection and has been downloaded from the Kaggle. The same dataset and two more are used for three types of infrastructures such as walls, pavements and decks in the previous work, but here in our model we used only one dataset of walls and apply the model and then achieved our desired output. First, we access the dataset form the drive and then unzip the dataset and inflated all the forty thousand images of walls including cracks and non cracks. By importing all the important and required libraries and packages, we loaded the dataset by making the directories for positive and negative datasets. We checked all the images in the datasets by defining them as a crack images and no crack images and make a function for the finding and counting their numbers of images in each dataset. Twenty thousand images are found in each of dataset. A module ImageDataGenerator is used and applied dataset dividing in the validation set and in the training set. Set the image height and width is same such as 128, also define the batch size equal to 32. The 32-batch size is a better size for the data of images during the training of model. We made three dataset names through coding name as train dataset, validation dataset and the test dataset. The training images are belonging to two classes, also there have been found and calculating 38 thousand images in the training, 40 thousand images in the testing test and 2 thousand images in the validation set by setting the keras ImageDataGenerator class and gave only 0.05 percent to the validation data. It means that, we give only 2 thousand images to the validation set by first evaluating the model and predict its output. In the validation set and in the testing set also found two classes to which all the images belong to. There are found form the training set that the training set has only a binary class classification because the result shows only negative compare with 0 and the positive compare with 1. After making the data ready, we applied the convolutional neural network (CNN) and made the CNN model for data and for feature extraction. The main feature extractor layers in the convolutional neural model are the convolution layer and the MaxPooling layer. We added two layers of convolution and two layers of maxpooling to the convolutional neural network (CNN). In CNN, first we defined input shape of the input images by calling the input function form tf.keras. define the image height, image width and the color channel for all the images. In the first layer of the CNN, the convolution 2 dimensional and the Maxpooling 2 dimensional are applied. The kernel size is set as 3x3 and then added second layer to the CNN. The second layer of the CNN also contains the same two layer such as convolution and the Maxpooling layer. Then applied the global average pooling layer and Dense layer to the final layer. We create the model by the name as model and passes the arguments in the model function name as inputs and outputs. The model summary is shown in the figure by showing its total parameters, trainable parameters and non-trainable parameters. There are 5000+ parameters, all are the trainable parameters. A parameters are the variables which are used during the training process. All the parameters are variables, that are learned during the process. The training of a model is done for purpose of checking its accuracy. We set 20 epochs to learn the patterns or features inside all the images. The epochs are just setting a parameter to repeatedly check the images and find more and more features inside the data during the training. After the completion of 8 epochs, the model is extracted and learned all the patterns and extracted the cracks very well after the 8 epochs. The validation accuracy is achieved 99%. For the first epochs, the ETA is 173 seconds. The Estimated time atteration for the last epochs is 142 seconds. We also evaluate the test dataset by finding the test accuracy. The accuracy of the test dataset is 98 percent. We got 9000+ correct predictions out of 10,000 records in the test set. The correct predictions of confusion matrix are shown in the figure 2.

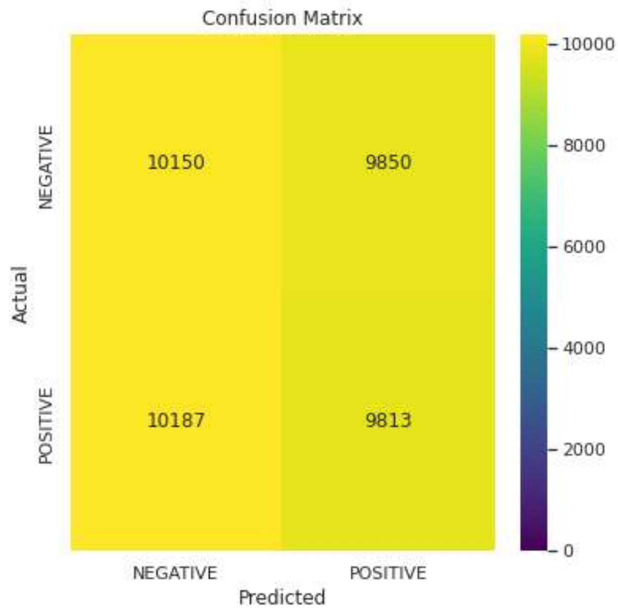
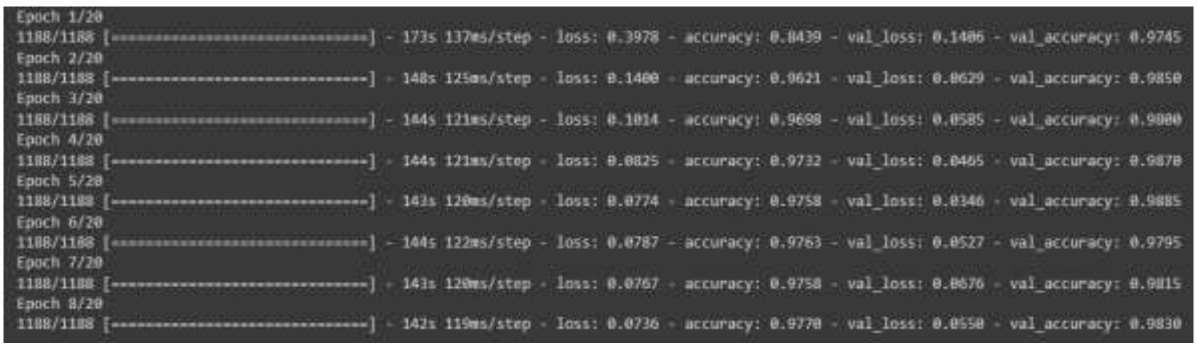


Figure 2: Shows correct predictions of all the input images.



```
Epoch 1/20
1188/1188 [==============================] - 173s 137ms/step - loss: 0.3978 - accuracy: 0.8439 - val_loss: 0.1406 - val_accuracy: 0.9745
Epoch 2/20
1188/1188 [==============================] - 148s 125ms/step - loss: 0.1400 - accuracy: 0.9621 - val_loss: 0.0629 - val_accuracy: 0.9850
Epoch 3/20
1188/1188 [==============================] - 144s 121ms/step - loss: 0.1014 - accuracy: 0.9698 - val_loss: 0.0585 - val_accuracy: 0.9800
Epoch 4/20
1188/1188 [==============================] - 144s 121ms/step - loss: 0.0825 - accuracy: 0.9732 - val_loss: 0.0465 - val_accuracy: 0.9870
Epoch 5/20
1188/1188 [==============================] - 143s 120ms/step - loss: 0.0774 - accuracy: 0.9758 - val_loss: 0.0346 - val_accuracy: 0.9885
Epoch 6/20
1188/1188 [==============================] - 144s 122ms/step - loss: 0.0787 - accuracy: 0.9763 - val_loss: 0.0527 - val_accuracy: 0.9795
Epoch 7/20
1188/1188 [==============================] - 143s 120ms/step - loss: 0.0767 - accuracy: 0.9758 - val_loss: 0.0676 - val_accuracy: 0.9815
Epoch 8/20
1188/1188 [==============================] - 142s 119ms/step - loss: 0.0736 - accuracy: 0.9770 - val_loss: 0.0550 - val_accuracy: 0.9830
```

Fig. 3: This is the overall accuracy report of the model, its val-accuracy and training accuracy, val-loss and training loss

Result and discussion

The present study utilizes 2,000 images for the validation set, 38,000 images for the training set and 40,000 images for the testing set just for the extraction of features (crack detection) from the whole model. The CNN is done a good work and give the spectacular result after being the trained the model and after the extraction of features such as cracks detection and no cracks. The validation accuracy is achieved in this model is 99 % and the least validation loss for the last epochs is 0.0550. this result in the validation loss shows that very less loss is occurred in the validation set, the model is learned their patterns in the images very well. Also, the training set accuracy is achieved is 98 percent and loss in the training set is 0.0736, this result shows very less amount of loss is happen in the training set. The final accuracy of the test set is 98.01% and the loss in the test set is 0.06720. The overall model is achieved a spectacular result by learning and finding the cracks and no cracks in any of the custom image. We sets the probability of cracks for any of the image is 0.5 %. If there are detected cracks in the image above 50%, the result will be a detected cracks other wise no cracks will be printed. It gives the same result, and the model can able to find cracks, when there have been present cracks in the image. The input images that are given to the model for detecting cracks are shown in the figures.

Fig. 4: input image for prediction

In this input image, the cracks are looking very clearly, this a normal image which we have taken from the ordinary wall of the house. The cracks are happened between the roof the house and the wall of the house.

Fig. 5 input image for prediction

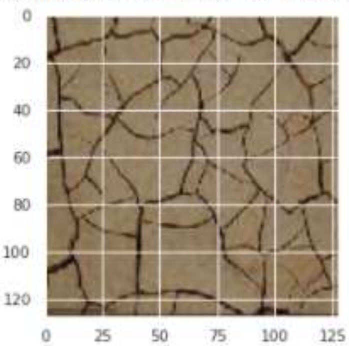
This is also the input image which we have been taken from the wall and very looking very simple. There are no cracks in the image, just the reflection of light is seen in the image.

Fig. 6 input image prediction

We have been given our own custom images to the trained model. The first three images are the input images on which a model is making its prediction. The convolutional neural network (CNN) is applied which is the class of the deep neural network. The convolutional neural network is given the best result by predicting the probability of any input images which has been applied to the model. The model is trained and can able to make prediction on any images.

The output predictions are shown in the figures below.

```
[ ] plt.imshow(img)

<matplotlib.image.AxesImage at 0x7fbb329e0150>
```




```
[ ] img1=image.img_to_array(img)
    img1=img1/255
```

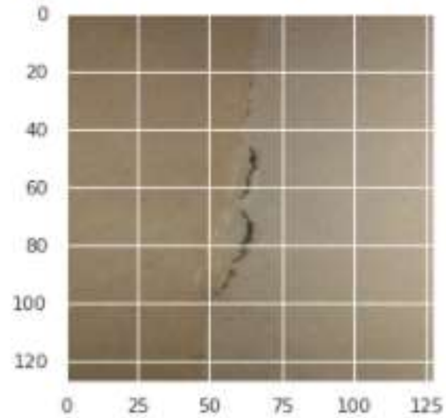
```
[ ] img1=np.expand_dims(img1,[0])
    print(img1.shape)

(1, 128, 128, 3)
```

```
[ ] #prediction 1
    prediction = model.predict(img1)

    print(prediction[0][0])
    display_prediction_stats(prediction[0][0])

1.0
Probability that the sample has a Crack = 1.0
Crack Detected !
```

Fig. 7: prediction of 1st input image

This result shows that the model is learned and well trained based on the cracks on the crack detection. In this image 1.0 is the probability of the sample of image which has a crack and are detected 100 percent. 1.0 percent cracks mean that 100 percent cracks are detected in the input image. This image is shown a good well probability of the prediction of the model. In this input image lots of cracks are looking and are very clearly seen by the eyes. One can look to this image, visually it will be decided that this image has full of cracks. That's why the model is also decided their output prediction on the image and extracted 100 percent cracks.

```
[92] plt.imshow(img)

<matplotlib.image.AxesImage at 0x7f89e07a4c10>
```




```
img1=image.img_to_array(img)
img1=img1/255

[94] img1=np.expand_dims(img1,[0])
     print(img1.shape)

(1, 128, 128, 3)

[95] #prediction 1
     prediction = model.predict(img1)

     print(prediction[0][0])
     display_prediction_stats(prediction[0][0])

0.5103202
Probability that the sample has a Crack = 0.5103202
Crack Detected !
```

Fig 8: Output prediction of 2nd input image

This image is also captured on the cell phone. This is the image which contains a crack but not more cracks like the first input. The convolutional neural network is extracted some the cracks and decided that, in this output prediction image has a crack and the probability of the cracks is 0.51 percent. This means that, the probability of the cracks is above the 50 percent that's why the result that in this output prediction sample of the image has some cracks.

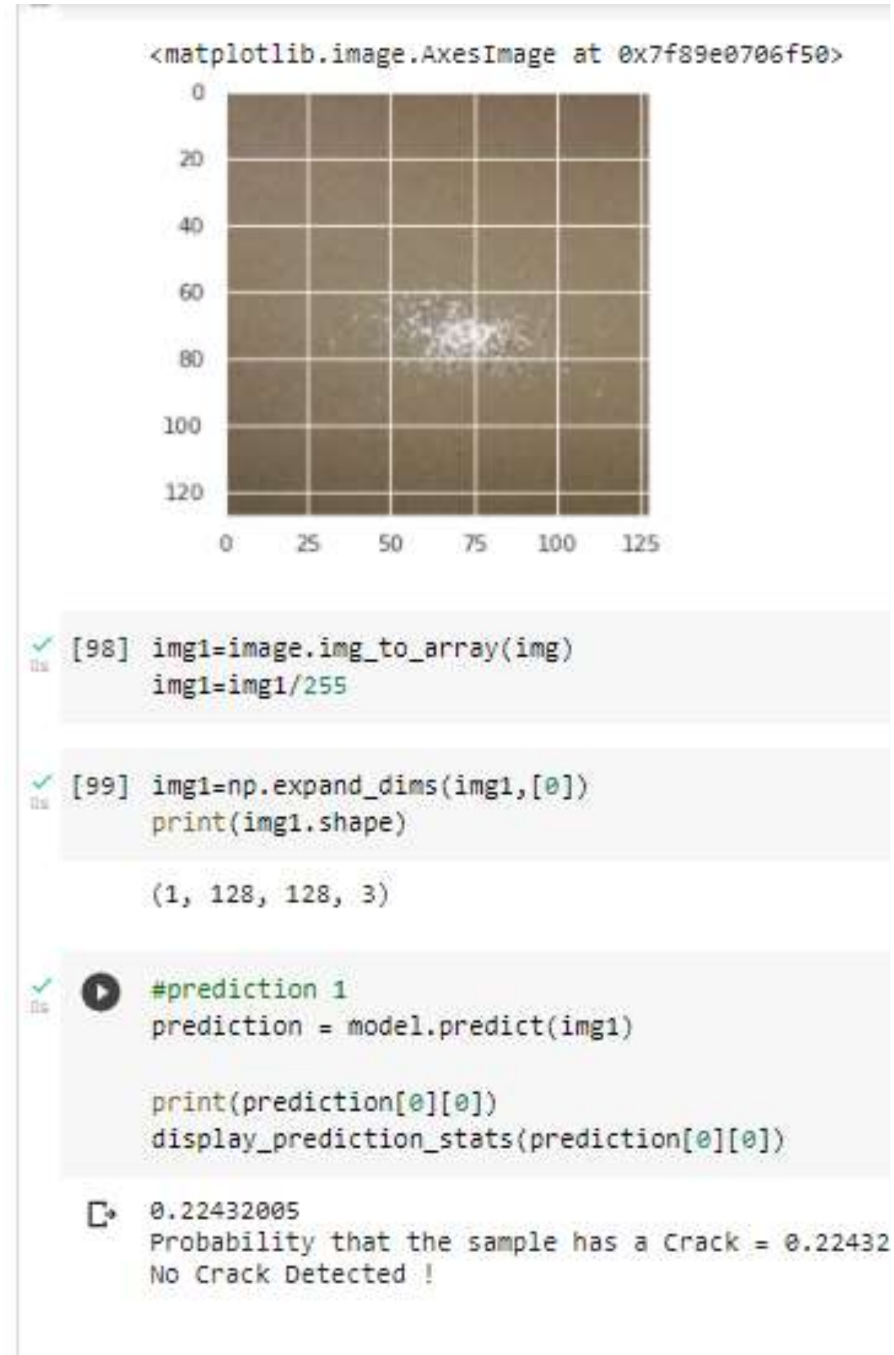


Fig. 9: output prediction of 3rd input image

This image has been captured form the wall and the cement of the wall is look like in good condition. In this image, no clearly cracks are visually looking. So, the model is predicting the cracks in the image but are less in amount. So, the probability of the cracks is below the threshold. The result shows no cracks in the given image.

Conclusion

Cracks are a critical sign of damage that happens in infrastructure. Such cracks indicate that monitoring and inspection are highly needed over time. Such as traditional methods of inspecting buildings and infrastructure cracks are very time-consuming, faulty, and challenging to assess, as well as methods to resolve the issues. Thus, the present study and the training of a model are based on the computer vision-based recognition and identification of cracks within the image of the walls. The dataset of the images contains two types of images. One dataset is named 'positive' and the other is named 'negative'. The positive dataset contains an image with cracks, whereas the other dataset does not have any cracks in its images. The convolutional neural network (CNN) is applied and extracts more features from the images. There are two main feature extractors in the CNN, which are the convolution layer and max pooling layer.

The result of the applied method demonstrates that it is a straightforward approach to training a model, yielding excellent results by extracting features from the input images. The accuracy of the predicted images is 100% in the first image, 51% in the second image, and 22% in the last image. Based on the image and its cracks, the model predicts its cracks. Regarding crack detection accuracy, the applied method can be further improved to achieve better accuracy, particularly when additional work is done in the CNN layers. If we increase the layers of the convolutional neural network, it will give us a more accurate result.

References

[1] M. M. "Machine learning techniques for structural health monitoring of heritage buildings: A state-of-the-art review and case studies," *Journal of Cultural Heritage,* p. 19, 2020.

[2] H. D. Wehle, " Machine learning, deep learning, and ai: What's the difference," *In Internationan Conference on Data scientist innovation day, Bruxelles, Belgium.,* pp. 1-6, July 2017.

[3] N.-D. Hoang, "Image processing-based recognition of wall defects using machine learning approaches and steerable filters," *Computational intelligence and neuroscience,,* pp. 1-19, 2018.

[4] C. and M. , "Automatic Pavement Crack Detection based on Image Recognition," *International Conference on Smart Infrastructure and Construction 2019 (ICSIC): Driving data-informed decision-making* , pp. 361-369, 2019.

[5] Q. and Z., "The method of the road surface crack detection by the improved otsu threshold.," *In 2019 IEEE International Conference on Mechatronics and Automation (ICMA),* pp. 1615-1620, 2019.

[6] C. V. Dung and L. D. Anh, "Autonomous concrete crack detection using deep fully convolutional neural network, Automation in Construction," vol. Volume 99, pp. 52-58, 2019.

[7] S. c. and B. K., "Image-based concrete crack assessment using mask and region-based convolutional neural network," *Structural Control and Health Monitoring,,* vol. Volume 26, no. Issue 8, 2019.

[8] S. R. F. M. Nejad and H., "An image-based system for pavement crack evaluation using transfer learning and wavelet transform.," *International Journal of Pavement Research and Technology,* vol. 14, no. 4, pp. 437-449, 2021.

[9] P. V. K. M. and G. V. , "Analysis of cracks in structures and buildings.," *In Journal of Physics: Conference Series,* vol. vol. 1706, no. No. 1, pp. 1-10, 2020.

[10] N. . J. M. and T. , "Survey on railway wheel defect detection using machine learning.," *Aut Aut Research Journal,* vol. 11, no. 4, pp. 1-12, 2020.

[11] M. M. Ata, K. M. E. and M. A. M. , "Toward Palmprint Recognition Methodology Based Machine Learning Techniques," *EJECE, European Journal of Electrical Engineering and Computer Science,* vol. Vol. 4, no. No. 4, pp. 1-10, 2020.

[12] A. K. "Machine learning applications in cell image analysis," *Immunology and cell biology,* vol. 95, no. 6, pp. 525-530, 2017.

[13] Y. L. Y. . B. and G. H., "Deep learning," *nature,* vol. VOL 521, pp. 436-444, 28 MAY 2015.

[14] R. Y. M. N. R. K. G. D. and K. T., "Convolutional neural networks: an overview and application in radiology.," *Insights into imaging,* vol. 9, no. 6, pp. 611-629, 2018.