

## Investigation and Analysis of Crack Detection using UAV and CNN: A Case Study of Hospital Raja Permaisuri Bainun

Goh Wei Sheng<sup>1</sup>, Wan Isni Sofiah Binti Wan Din<sup>2,\*</sup>, Quadri Waseem<sup>2</sup>, and Azlee Bin Zabidi<sup>2</sup>

<sup>1</sup>Intel Microelectronics (M) Sdn.Bhd, Halaman Kampung Jawa, Kawasan Perindustrian Bayan Lepas, 11900 Bayan Lepas, Pulau Pinang, Malaysia.

<sup>2</sup>Faculty of Computing, Universiti Malaysia Pahang, 26600 Pahang, Malaysia.

**ABSTRACT** – Crack detection in old buildings has been shown to be inefficient, with many technical challenges such as physical inspection and difficult measurements. It is important to have an automatic, fast visual inspection of these building components to detect cracks by evaluating their conditions (impact) and the level of their risk. Unmanned Aerial Vehicles (UAV) can automate, avoid visual inspection, and avoid other physical check-ups of these buildings. Automated crack detection using Machine Learning Algorithms (MLA), especially a Conventional Neural Network (CNN), along with an Unmanned Aerial Vehicle (UAV), can be effective and both can efficiently work together to detect the cracks in buildings using image processing techniques. The purpose of this research project is to evaluate currently available crack detection systems and to develop an automated crack detection system using Aggregate Channel Features (ACF) that can be used with unmanned aerial vehicles (UAV). Therefore, we conducted a real-world experiment of crack detection at Hospital Raja Permaisuri Bainun using DJI Mavic Air (Drone Hardware) and DJI GO 4(Drone Software) using CNN through MATLAB software with CNN-SVM method with the accuracy rate of 3.0 percent increased from 82.94% to 85.94%. in comparison with other ML algorithms like CNN Random Forest (RF), Support Vector Machine (SVM) and Artificial Neural Network (ANN).

### ARTICLE HISTORY

Received: 25 August 2022

Revised: 5 November 2022

Accepted: 16 November 2022

Published: 1 January 2023

### KEYWORDS

*Crack Detection*

*Structures*

*Unmanned Aerial Vehicle*

*Machine Learning*

## INTRODUCTION

Concrete cracks are the major concern of all old buildings (structures). These cracks are created due to their old age or due to various other environmental factors. These cracks are very common in old buildings. Cracks are the major issues that need to be addressed to guarantee the safety, serviceability, and robustness of any architecture [1]. Some of these cracks might be very dangerous and can't be ignored for safety reasons. It would be very unrealistic to expect a crack-free old building. Therefore, crack detection for such buildings is a must to locate the cracks earlier to prevent the occurrence of any future serious issues. The reasoning is that when cracks begin to form and spread, it can cause a reduction in the effective loading area, resulting in stress expansion and subsequent failure of the concrete or other structures [2]. Due to limitations in strengthening concrete structures and buildings, deterioration over time and cracking seem unpreventable and occur in all forms of structures, such as concrete walls, slabs, and beams. Especially for concrete components, cracks produce a gap that is harmful and allows corrosive chemicals to penetrate through the structure [3], which weakens the strength of these buildings. As a result, rapid and reliable surface crack detection and analysis using automated procedures is critical to replacing human inspectors, who are slow and prone to error [7]. Recent reviews [6,7,9,10] identify the rise of the trend of implementing the technique of image processing to speed up the efficiency of crack detection in structures. It shows that it is essential to evaluate the visual condition of vertical and horizontal structural elements. The details of cracks can be analysed and determining the suitable recovery methods to repair the damaged structures and avoid crisis failures becomes a need [11]. Fortunately, machine learning and unmanned aerial vehicles (UAV) are playing a key role in the industry of the Internet of Things (IoT), in fact, the Internet of Everything (IoE). Hence, we can try and use these UAVs along with image processing related ML to enhance and improve the performance of the crack detection methods due to fact that the Machine learning can provide the best results for the data/pictures collected from UAV source [15,16,17,18,19,20].

The objective of this research is to investigate and perform a depth analysis of the latest crack detection techniques using Unmanned Aerial Vehicles (UcAV) and Machine Learning algorithms (MLA) especially CNN-SVM algorithm and compare our results with other ML algorithms, which are related to our research project. Convolutional neural networks (CNNs) are used to detect crack in images to do away with the extraction of crack features. This research aims to determine the way to analyse crack detection at Hospital Raja Permaisuri Bainun using an Unmanned Aerial Vehicle (UAV) for UEM Edgenta using Aggregate Channel Features (ACF). Therefore, a conventional neural network (CNN) algorithm was developed for the analysis of crack detection which evaluated the impact and the level of risk caused by their cracks.

## RELATED WORKS

Prior to the development of automated/ML based crack detection techniques integrated with UAV, most of the technique involve physical checking and data collection with a time and labour constraints. A speedier method based on Machine learning algorithms and UAV was utilized to avoid the previous drawbacks. As a conventional method for crack detection, image binarization, which is often adopted for text recognition and medical image processing [4], was used as a solution for crack detection. This is because there was a similar characteristic between the texts and the cracks in the form of lines and curves. But when we consider the technical hindrances versus performance in the long run, it simply fails to qualify.

Another standard conventional method detection approach, which was adopted by the Otsu method [5], is also unsatisfactory. This is because crack detection undergoes challenges like uneven illumination, low contrast, the existence of shading, noise pollution, blemishes, or concrete spall in images [6]. These challenges will cause inaccuracy in the image binarization approach, which relies on the quality of the image, the identity of the surface of the background, and related parameters [7]. The improvements towards crack detection that rely on the image binarizing approach are continuously studied within the researcher community. Moreover, these conventional methods of crack detection for buildings have been proven to face many difficulties. Nevertheless, most developing countries are still manually doing crack detection tasks. Thus, a longer period and more effort are required to detect, collect data, and analyse the cracks. Both [8] and [9] stated that it is very important to have a visual examination of the structural components to discover cracks and evaluate the conditions of the building. Another works are utilizing the concept of crack detection using latest machine learning algorithms via different forms [12,13,14]. The comparison and evaluation of those main three related algorithms with our proposed algorithm is done in next section.

## COMPARISON OF EXISTING WORKS

In this section, we have evaluated three different algorithms with our proposed algorithm. The evaluation results have proven remarkable accuracy and perfection in detecting the cracks in our use case using our combination CNN algorithm with UAV assistance as compared to their opponent algorithms.

### Bridge Crack Detection using Multi-Rotary UAV and Object-Base Image Analysis:

In this research [12], there are two major targets for inspection, which are cracks and concrete spalling. The method describes how automatic crack detection works and how to figure out the volume of concrete spalling. The technique illustrates the outcome of each pre-processing step and analyses the crack detection results. Figure 1 is an image of the abutment, with cracks passing through it almost horizontally. Due to many non-cracked objects on the abutment surface (e.g., cables, stains, pipes), the image is very complex. Moreover, due to insufficient light, the original image appears darker. The process of linear stretching can enhance the overall contrast. But the upper right corner is darker than other areas due to the light. The darker side had been blocked by the viaduct. As shown in Figure 2, this uneven illumination can be significantly reduced by rolling ball background subtraction.

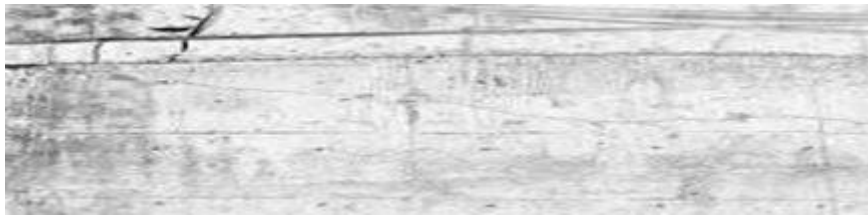


Figure 1. Crack Image.



Figure 2. The result after background subtracts.

After the process of subtracting the background, it is not easy to detect some tiny and unclear cracks. The researcher enhances these cracks by using a local contrast enhancement technique. The last stage is to use object-based image analysis (OBIA) technology to develop a regulation set for crack detection. The pre-processed image will be divided into

several unique image objects. Then, we use these unique object characteristics and classified regulation sets to assort them into the crack and non-crack groups.

### Detection of Asphalt Pavement Potholes and Cracks Based on the Unmanned Aerial Vehicle Multispectral Imagery

Asphalt road pavement is nearly the most common type of road surface. However, the pavement appears to be the combined effect of the combined effects of ageing and road surface deterioration. Two general types of road surface damage that will cause a major impact on vehicles are cracks and potholes [13]. There are two main methods for digital pavement images to extract the surface defects, which are image processing and machine learning algorithms [10]. In this research, the researchers applied the Support Vector Machine, Artificial Neural Network and Random Forest (SVM, ANN, and RF) classifiers to identifying the asphalt road surface cracks and potholes from the images collected. The collected sample data was then used for training and validation. The training details are presented in Figures 3 and 4.

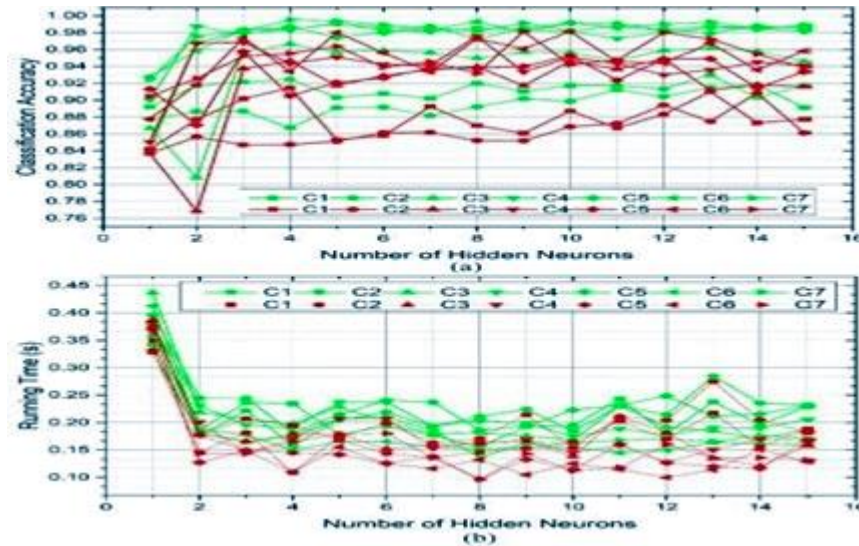


Figure 3. ANN algorithms training.

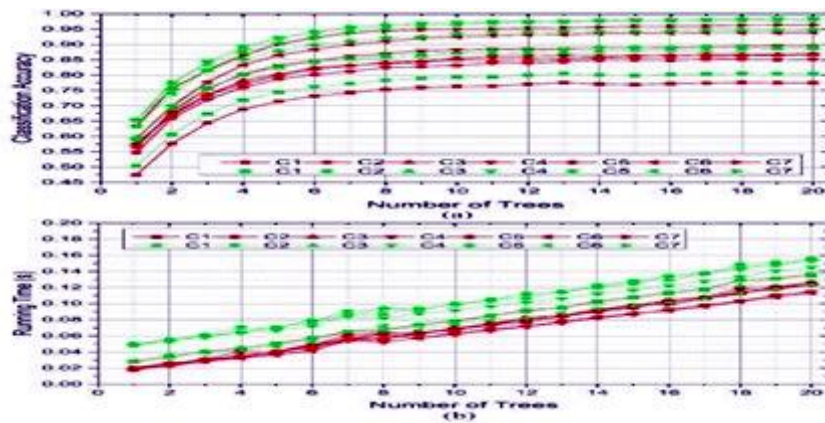


Figure 4. RF Algorithms training.

Table 1. Result of Training.

| Models          | SVM    | ANN    | RF     |
|-----------------|--------|--------|--------|
| Accuracy        | 98.97% | 98.46% | 98.43% |
| Running Time(s) | 0.63s  | 0.21s  | 0.09s  |

The figures (Figure 3 and Figure 4) and table (Table 1) show the classification accuracy and running time of each classifier, but the results might differ in different data sets and different applications. It is concluded that the accuracy performance and running time of each classifier depend on the data sets and application field.

### Crack Detection in Masonry Structures using Convolutional Neural Networks and Support Vector Machines:

Research in [14] is mainly focusing on proposing crack detections that merge the CNN and SVM methods. Here, CNN is used to extract the characteristics from RGB images and SVM is used as another option in the classifier for the softmax layer to enhance classification capabilities. A digital camera and a UAV are used to collect a dataset containing images of cracks from the sites. A series of images collected are used for training and validation. The result proves that a combination of the CNN and SVM models could have better performance than the CNN model alone. The system can also automatically detect the images of masonry structures, bringing convenience for the inspection of structures.

This proposed system consists of three modules. To begin, the images are captured using a DSLR (digital single-lens reflex) camera and a UAV. Then, the crack detection system will be classified. The results are generated to be used in the final module. These modules are mentioned in the Figures 5 and 6. In this last step, which is image acquisition, the DJI Phantom 4 drone was used to capture the images. The drone was programmed to fly at two different heights. In addition, there are more samples of images that were collected as shown in Figures 7 and 8.

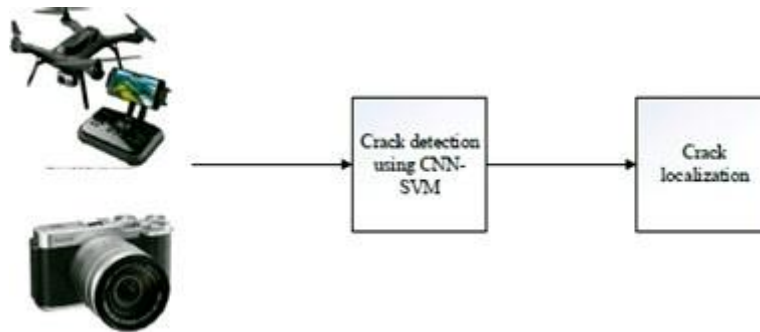


Figure 5. The outline of the proposed methodology.



Figure 6. The outline of the proposed crack detection system.



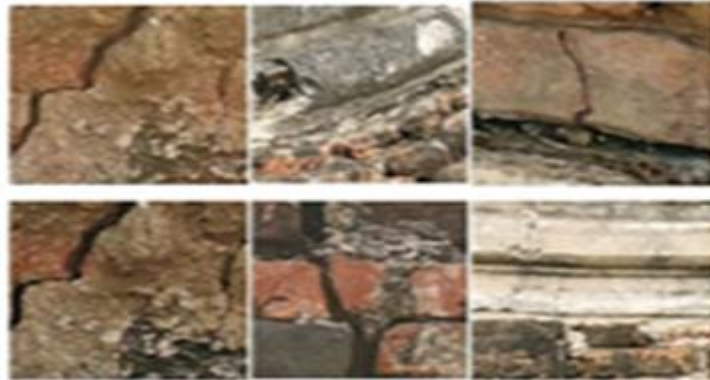
Figure 7. Sample images acquired using a UAV.





**Figure 8.** Samples of images of the Chapel viaduct.

The images collected were trained and classified by CNN using a MATLAB algorithm. In the training, the patches that belonged to the masonry areas were selected to train, while the patches involving surrounding objects such as trees were ignored. In this work, there are a total of 6002 image patches. Image patches used were classified into cracks and non-cracks images. While 3162 patches of 6002 were trained and validated, and the left patches were tested. The crack patches are labelled as 0, while the non-crack patches are labelled as 1. Examples of patches are presented in Figures 9 and 10.



**Figure 9.** An example of crack patches.



**Figure 10.** An example of non-crack patches.

The second step is crack detection. From the proposed system, CNN has been applied because it can solve a lot of real-world problems rapidly. In the CNN architecture, there are two main tools, which are a multilevel deep feature extractor and a classifier. The multilevel deep feature extractor is used as a tool to get the distinguishing characteristics from image pixel intensity values, while the SVM classifier is for the classified uses. The last step is to localise cracks in the final module. The validation and testing datasets are conducted on the proposed system to do the evaluation. The evaluation is done on the system to predict inaccuracies that would happen. Evaluation of the performance is done by using the Receiver ROC analysis, confusion matrix, and classification report. The technique of cross-validation is used to gain the most ideal values due to the Radial Basis Function (RBF), which is used as a kernel. Table 2 shows the results of the parametric study for SVM, and Table 3 shows the confusion matrix for class classification.

**Table 2.** Parametric Study for SVM.

| C | Gamma | Accuracy |
|---|-------|----------|
| 1 | 0.5   | 0.770    |
| 1 | 1     | 0.72     |
| 2 | 1     | 0.72     |
| 3 | 1     | 0.73     |
| 4 | 1     | 0.73     |
| 5 | 1     | 0.71     |

**Table 3.** Confusion Matrix for Class Classification.

| Ground Truth Label   | Predicted label     |                      |
|----------------------|---------------------|----------------------|
|                      | Positive Crack      | Negative (Non-Crack) |
| Positive (Crack)     | True Positive (TP)  | False Negative (FN)  |
| Negative (Non-Crack) | False Positive (FP) | True Negative (TN)   |

The following five equations are used for the classification analysis in the classification report.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$precision = \frac{TP}{TP + FP} \quad (2)$$

$$TPR = Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = \frac{2 \times precision \times Recall}{precision + Recall} \quad (4)$$

$$FPR = \frac{FP}{FP + TN} \quad (5)$$

In this research, an experiment has been conducted to compare the results of CNN alone and the proposed system. Table 4 determines that the CNN-SVM method has an accuracy rate of 85.94% while the CNN method only has 82.94%. The CNN-SVM method is also shown to be better in other metrics. The results of the validation and testing dataset are shown in Table 4 and Table 5. Figure 11 is a plot between TPR and FPR (True and False positive Rate) for the possible values of the output as computed by comparing predicted labels to ground truth values. The Equation 3 and Equation 5 are for TPR and FPR calculations used respectively. Figure 11 shows that the CNN-SVM method is more towards the top left corner of the graph, so we conclude that CNN-the SVM method is much better than the CNN method. Sample images are used to localise the cracks to gain the result. The image is divided into several grids, and each of the grids was classified by the system. The crack localization is shown in Figure 12.

**Table 4.** Results of The Validation Dataset.

| Method  | Validation Accuracy | Precision | Recall | FI Score |
|---------|---------------------|-----------|--------|----------|
| CNN     | 82.94               | 0.83      | 0.71   | 0.74     |
| CNN-SVM | 85.94               | 0.84      | 0.79   | 0.79     |

**Table 5.** Results of The Testing Dataset.

| Method  | Accuracy | Precision | Recall | F1-Score |
|---------|----------|-----------|--------|----------|
| CNN     | 67.5     | 0.80      | 0.68   | 0.73     |
| CNN-SVM | 74.9     | 0.82      | 0.78   | 0.78     |

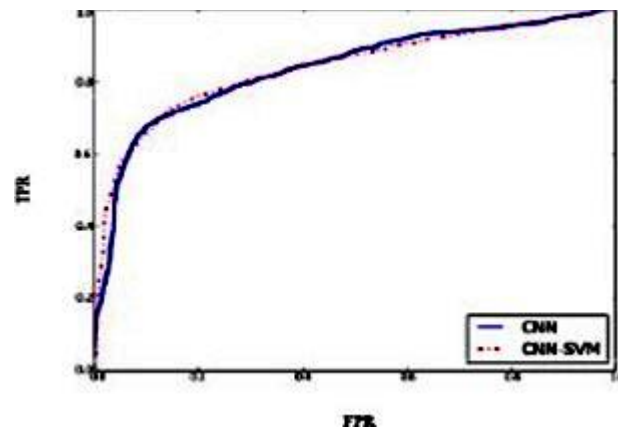


Figure 11. The ROC Curves of CNN-SVM and CNN model.

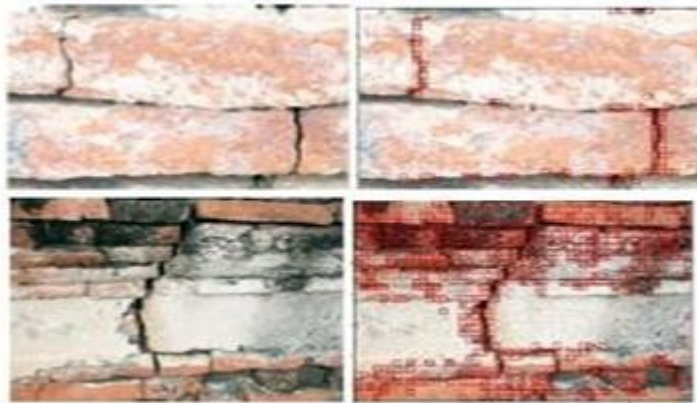


Figure 12. Crack localization of sample images.

### Motivation Derived from Comparison of Existing Works

From Performance Evaluation section, it was observed their three sub-sections (a, b, and c) have a difference in goals and objectives. Sub-section a: aims to propose a bridge crack detection system using UAV and OBIA (Object-based image analysis). Sub-section b: proposed the detection of asphalt pavement potholes and cracks based on UAV multispectral imagery. Sub-section c: proposed crack detection in masonry structures using CNN and SVM. Thus, this proposed research needs the elements from sub-sections a, b, and c. So, section a gives idea to be careful about things like shadows, uneven lighting, and dirt on the building, which can have an effect. Subsection b proposes that the accuracy and running time of SVM, ANN, and RF might be affected by the dataset or application environment, so any of the classifiers could be efficient depending on the research dataset and application environment. For example, Subsection c proposes that the classifier can not only be independent; they could be a mix. For example, the CNN mixed with the SVM to form CNN-SVM for higher recognition accuracy (ratio of correctly identified crack and non-crack patches to the total number of input patches) and better results.

Hence, we conclude that neural networks, especially conventional neural networks (CNN) would be used in our research due to their various benefits and with the performance comparison, it can provide the basic target features as per our requirements in this research project by utilising the (a, b, and c) of Section of Performance Comparison Investigation. From the above studied that CNN, SVM, ANN, RF, CNN-SVM detectors are used for detection purpose. In this research work, Aggregate Channel Features (ACF) detector would be used for crack detection purpose.

### MATERIAL AND METHODS

Machine Learning Algorithms (MLA), particularly a Conventional Neural Network (CNN), and an Unmanned Aerial Vehicle (UAV), are developed to work well together for automatic detection of cracks in buildings. CNNs are deep learning algorithms that were created from ANNs, and they excel at object and picture classification. Because of their partial connections, ability to share weights, and neuronal "pooling" process, CNNs can learn picture information more quickly than ANNs by requiring fewer calculations per parameter and hence, can provide the basis, platform, and generate ideas related to our case study for efficient outcomes.

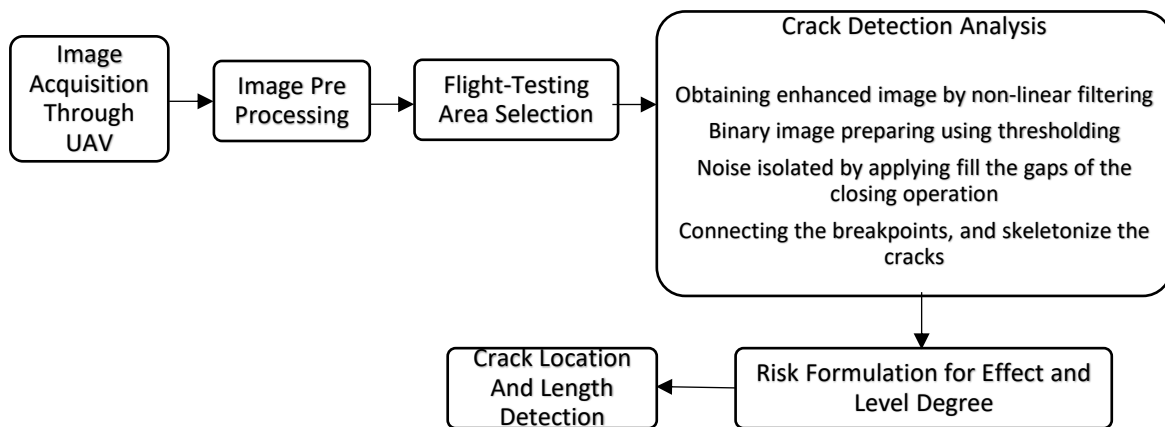
We evaluate the available crack detection technologies and develop an automated crack detection system that can be used with unmanned aerial vehicles (UAV) for our case study of Hospital Raja Permaisuri Bainun in Malaysia. To detect the cracks, we conducted a real-world experiment at Hospital Raja Permaisuri Bainun using DJI Mavic Air (drone hardware), DJI GO 4 (drone software), and CNN using MATLAB software and the CNN-SVM technique.

## RESEARCH METHODOLOGY

In this research project, the methodology opted is iterative waterfall methodology because it has the flexibility to allow going back to the previous stage when an error is detected, or if some troubleshooting is needed. It consists of repeating the five stages which are the Requirement Analysis stage, Design stage, Implementation stage, Verification stage, and Maintenance stage. Hospital Raja Permaisuri Bainun, originally known as Ipoh Hospital, Hospital Raja Permaisuri Bainun is a government hospital in Ipoh, Perak, Malaysia. It was built in 1891 as the District Hospital with a capacity of 50 beds and later promoted as the State Hospital in 1942, with a total capacity of 990 beds. In the year 1980, its 8-story main building was completed along with some other small buildings. This research project aims to determine the best way to analyse crack detection at Hospital Raja Permaisuri Bainun using an Unmanned Aerial Vehicle (UAV) for UEM Edgenta using Aggregate Channel Features (ACF) detector. Therefore, in the first phase, a conventional neural network (CNN) is developed by using MATLAB for analysis of crack detection and evaluating the impact and level of risk caused by these cracks.

### Experimental Setup

The software and hardware specifications include MATLAB and DJI GO 4, while hardware specifications include laptop and DJI MAVIC AIR. In this section, we present our data analysis along with experimental findings.



**Figure 13.** Block diagram for crack detection.

The block diagram for crack detection is presented in Figure 13 and the phase details are mentioned as:

#### Data Collection:

In the data collection phase, the Unmanned Aerial Vehicle (UAV) will collect the required data. The whole environment will be recorded throughout the real-world experiment. After finishing the recording, the footage will be brought back and proceed to the next phase, which is the analysis phase.

#### Data Analysis:

The collected data from the UAV is used for further processing. A CNN-based algorithm is developed and a flight-testing area is determined. This phase starts with determining a flight-testing area to conduct the real-world experiment. The real-world experiment was conducted at Hospital Raja Permaisuri Bainun using the data from the UAV. In this phase, the programme will be run on MATLAB. In its next step, it chooses the "Crack Detection Analysis" and inputs the pavement image. The analysis will be started by obtaining the enhanced image by non-linear filtering, followed by a binary image using thresholding. Both obtain the enhanced image, and the binary image will be applied to fill the gaps of the closing operation to eliminate the isolated noise. In its next step, the programme will connect the breakpoints, skeletonize the cracks, and eliminate the noise.

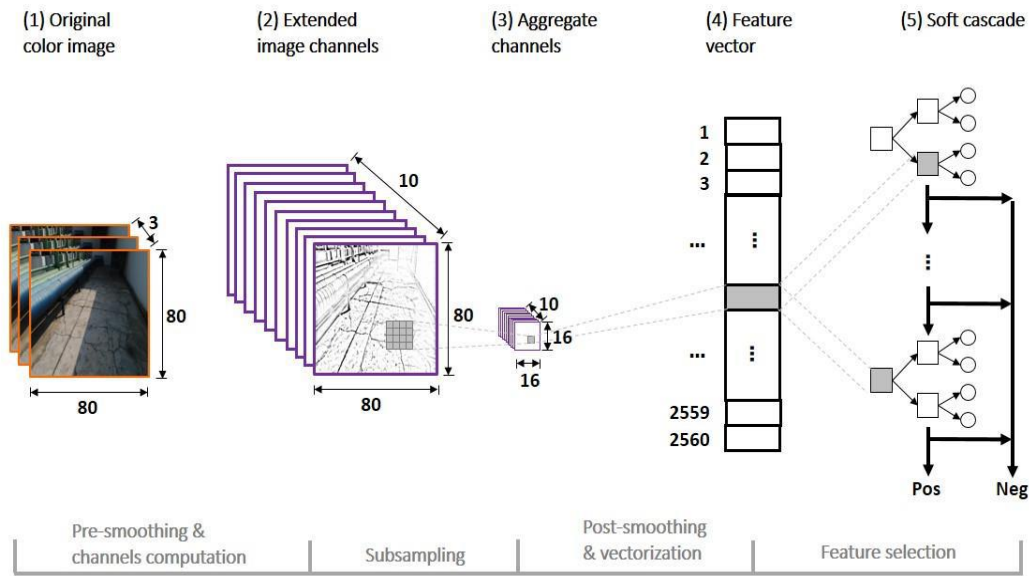
After finishing the analysis, the programme will show the result data. It will determine the crack location and length. The last step of our experiments would be to use the results to figure out the effect and level of risk. The expected outcome of the interface of our research project is mentioned in detail in Figure 14. Hence, it shows the processing and the outcome of the crack detection. Accurate and efficient results are produced.

#### Crack Detection Algorithm

In this section, we make a full exploration of the aggregate channel features in the context of crack detection. We first give a brief introduction of the feature itself, including its computation, properties and advantages over traditional Haar-like features used in VJ framework. Then the detailed experimental investigation is described in two parts, feature design and training design. Before that, some guidelines concerning how we conduct the investigation are demonstrated. Each design part is divided into several separate experiments ended with a summary explaining the specific parameters used in our proposed crack detector. Note that each experiment focuses on only one parameter and the others remain constant.



Through the well-designed experiments, the proposed crack detector based on aggregate channel features is built step by step.



**Figure 14.** Workflow of proposed crack detector

## Feature description

### Channel Extension:

The basic structure of the aggregate channel features is channel. The application of channel has a long history since digital images were invented. The most common type of channel should be the colour channels of the image, with gray-scale and RGB being typical ones. Besides colour channels, many different channel types have been invented to encode different types of information for more difficult problems. Generally, channels can be defined as a registered map of the original image, whose pixels are computed from corresponding patches of original pixels [21]. Different channels can be computed with linear or non-linear transformation of the original image. To allow for sliding window detection, the transformations are constrained to be translationally invariant.

### Feature Computation:

Based on the definition of channels, the computation of aggregate channel features is quite simple. As shown in Figure 14, given a colour image, all defined channels are computed and subsampled by a preset factor. The aggregate pixels in all subsampled channels are then vectorized into a pixel look-up table. Note that an optional smoothing procedure can be done on each channel with a binomial filter both before computation and after subsampling.

### Classifier Learning:

The learning process is quite simple. Two changes are made compared with VJ framework. First is that weak classifier is changed from decision stump to depth-2 decision tree. The more complex weak classifier shows stronger ability in seeking the discriminant intra and inter channel correlations for classification. Second difference is that soft-cascade [21] structure is used. Unlike the attentional cascade structure in VJ framework which has several cascade stages, a single-stage classifier is trained on the whole training data and a threshold is then set after each weak classifier picked by Adaboost. These two changes lead to more efficient training and detection.

### Overall Superiority:

Compared with traditional Haarlike features used in VJ framework, aggregate channel features have the following differences and advantages:

- 1) The image channels are extended to more types to encode diverse information like color, gradients, local histograms and so on, therefore possess richer representation capacity.
- 2) Features are extracted directly as pixel values on downsampled channels rather than computing rectangular sums with various locations and scales using integral images, leading to a faster feature computation and smaller feature pool size for boosting learning. With the help of cascade structure, detection speed is accelerated more.
- 3) Due to its structure consistence with the overall image, when coupled with boosting method, the boosted classifier naturally encodes structured pattern information from large training data, which gives more accurate localization of cracks in the image.

## Investigation guidelines

All investigations are trained on the training data1 and tested on the testing data1. To make it clear, there are in total 4666 positive samples and 23330 negative samples selected from training data1 which are kept constant in all investigations. Testing data1 is used to conduct testing and in the testing video contains cracks that vary a lot in pattern, position and illumination. To alleviate the ground-truth offset caused by different annotation in training and testing set and make the evaluation more comparable, a lower Jaccard index<sup>3</sup> with threshold 0.05 is adopted in comparative evaluation. Practically the lower threshold won't cause errors being mistakenly corrected.

## Feature design

To fully exploit the power of aggregate channel features in crack detection domain, a deep investigation into the design of the feature is done mainly on channel types, window size, subsampling method and feature scale.

### Channel types:

Three types of channels are used, which are color channel (Gray-scale, RGB, HSV and LUV), gradient magnitude, and gradient histograms. The computation of the latter two channel types could be seen as a generalized version of HoG features. The Jaccard index is defined as the size of the intersection divided by the size of the union of the sample sets.

### Detection window size:

Detection window size is the scale to which we resize all crack and non-crack samples and then train our detector. Larger window size includes more pixels in feature pool and thus may improve the crack detection performance. On the other hand, too large window will miss some small crack and diminish the detection efficiency.

### Sub-sampling:

The factor for subsampling can be regarded as the perceptive scale for that it controls the scale at which the aggregation is done. Changing the factor from large to small leads to the feature representation shifting from coarse to fine and the feature pool size getting bigger.

### Smoothing:

As described in feature description, both pre and post smoothing is done in default setting of aggregate channel features. A binomial filter with a radius of 1 is used for smoothing. The smoothing procedure also has a great influence on the scale of the feature representation. Concretely, pre-smoothing determines how far the local neighborhood is in which local correlations are encoded before channel computation, while post-smoothing determines the neighborhood size in which the computed channel features are integrated with each other. In [22], the former corresponds to the 'local scale' of the feature, while the latter represents the 'integration scale'. We vary the filter radius used in pre and post smoothing and find that both using a radius of 1 gets the best results.

### Multi-scale:

In aggregate channel features, although hidden information at different scale could be extracted at a cost of more weak classifiers, it would be better to make the integrated channel features multi-scaled and thus make themselves more discriminant. Therefore, the same or better classification performance can be achieved with fewer weak classifiers. In this part, we implement three multi-scale version of aggregate channel features in the aforementioned three kinds of scale, perceptive scale (subsampling), local scale (pre-smoothing) and integration scale (post-smoothing) and compare their performances.

The color channel, gradient magnitude and gradient histograms prove themselves a good match in aggregate channel features. However, different choices of color channel used and on which gradients are computed have a great impact on performance. According to the experiments, LUV channel and gradient magnitude and 6-bin histograms computed on RGB color space (in total 10 channels) are the best choice for crack detection. Larger detection window size generally gets better performance but will miss many small faces in testing and lead to inefficient detection. In this work, we set the size to  $80 \times 80$  as its optimal performance. A subsampling factor of 5 is most reasonable according to the experiments, while different pooling methods show small differences. However, max pooling and stochastic pooling are much slower than average pooling, therefore the average pooling becomes the best match for the sake of efficiency. In this way, the resulting feature pool size of our crack detector is  $(80 \div 5) \times (80 \div 5) \times 10 = 2560$ , considerably smaller than that in VJ framework.

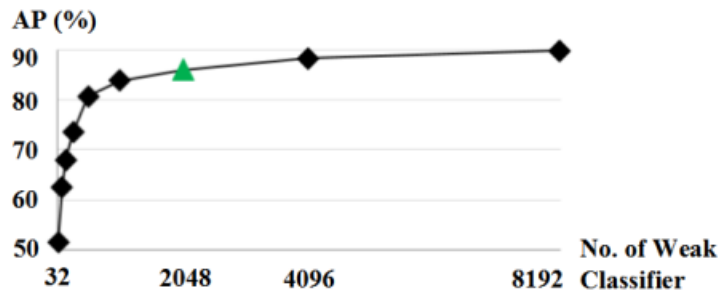
## Training design

Besides careful design of the aggregate channel features, experiments on the training process which is similar to that in VJ framework are also carried out. The differences are that the weak classifier is changed into depth-2 decision tree and soft-cascade [23] structure is used. Details of the training design are as follows.

### Number of weak classifiers:

Given a feature pool size of 4,000, we vary the number of weak classifiers contained in the soft-cascade. In Figure 15 performances of various numbers of weak classifiers ranging from 32 to 8192 are displayed, which shows that apparently more classifiers generate better performance, and when the number gets larger the performance begins to

saturate. Since more classifiers slow down the detection speed, there's a trade-off between accuracy and speed. Searching for the saturate point as the optimal is significant during training in such framework.



**Figure 15.** Comparison of different numbers of weak classifier in the soft cascade

### Training data

Empirically, more training data will get better performance given powerful representation capacity. In the training phase, the training data are used as the positive and negative training data. Based on observations above, we choose 2048 as the number of weak classifiers contained in the soft cascade. As each weak classifier is a depth-2 decision tree, it takes only two comparing operations to apply a weak classifier, which is quite fast.

### Overall Procedure to Analysis of Crack Detection

This project aims to determine the way to conduct analysis of crack detection at Hospital Raja Permaisuri Bainun for UEM Edgenta. The overall procedure to analysis of crack detection is shown in Figure 16. Therefore, it would start from determine a data collection area to conduct the real world experiment. The real world experiment is conduct at Hospital Raja Permaisuri Bainun using Unmanned Aerial Vehicle (UAV). The flight testing area that determined is on rooftop of the building of Hospital Raja Permaisuri Bainun. The whole environment will be recorded throughout the real world experiment. After finish recording, the collected video data will be bring back and proceed the next phase which is analysis phase.

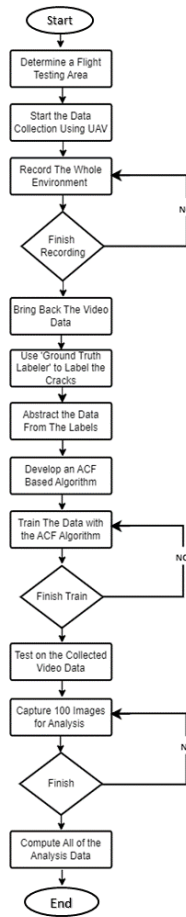
In the analysis phase, the collected data is used for further analysis. In this phase the crack in the collected data will be label using 'Ground Truth Labeler' on MATLAB. Cracks will be label in the 'Ground Truth Labeler' and the label will be named as 'Crack' and rectangle pixel label will be used to label the crack. After done labeling the data will be abstracted for training used. In the abstract process the label data will be abstract into 251 images and the images will be used for training the ACF detector. In the training phase, the images abstracted will be used. The training will take 10 stages with sampling factor of 5 and the model size is  $80 \times 80$ . In the training phase, 4666 positive samples and 23330 negative samples will be trained and the trained classifier has 2048 weak classifier. After the training done, a crack detection testing will be perform on the collected video data. In the crack detection testing, the trained ACF detector will automatic detect the crack from the testing data and used boundary box to box the crack detected. After the detection completed, analysis toward the performance of detection will be done. In the analysis process, 100 screenshot images will be selected randomly and analyze the percentage of accuracy. The last step would be collect and compute all the analyze data. The details are simplified through Figure 16.

### Procedure for Training Phase

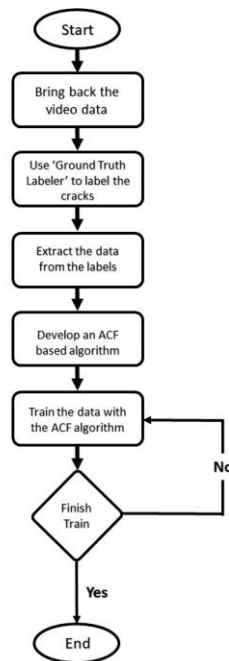
In the training phase, ACF detector is choose to develop the algorithm. The training phase start with abstract a part of the collected video data to train the detector. Ground Truth Labeler in MATLAB is used to label the cracks. After done labeling, the crack labels are abstracted to 251 images for training purpose. In the training phase, the images abstracted will be used. The training will take 10 stages with sampling factor of 5 and the model size is  $80 \times 80$ . In the training phase, 4666 positive samples and 23330 negative samples will be trained and the trained classifier has 2048 weak classifier. The details are simplified through Figure 17.

### Procedure for Analysis Performance of ACF Detector

After the training phase, the video data collected by Mavic Air is the input into the computer for analysis of crack detection. The procedure for analysis of crack detection is shown in Figure 18 below. The procedure start with open the crack detection program on MATLAB. Then, run the program for crack detection on the video data collected. In the crack detection testing, the trained ACF detector will automatic detect the crack from the testing data and used boundary box to box the crack detected. Next, screenshot 100 images randomly from the detection process. After that, do analysis based on the 100 selected images to determine the accuracy percentage of crack detection using ACF detector. Lastly, collect and compute all the analyze data. The details are simplified through Figure 18.

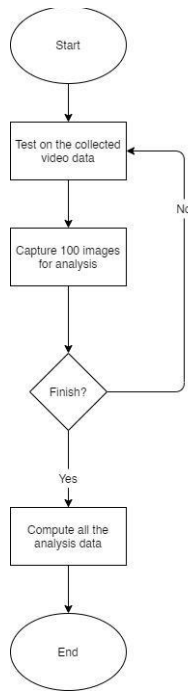


**Figure 16.** Overall Procedure to Analysis of Crack Detection

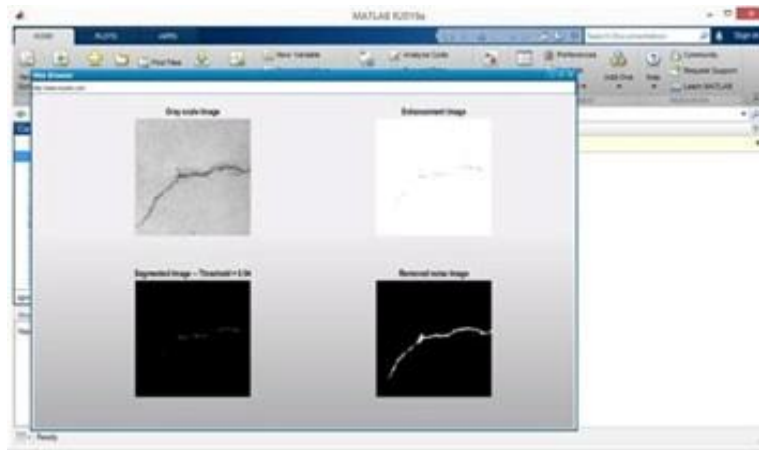


**Figure 17.** Procedure for Training Phase

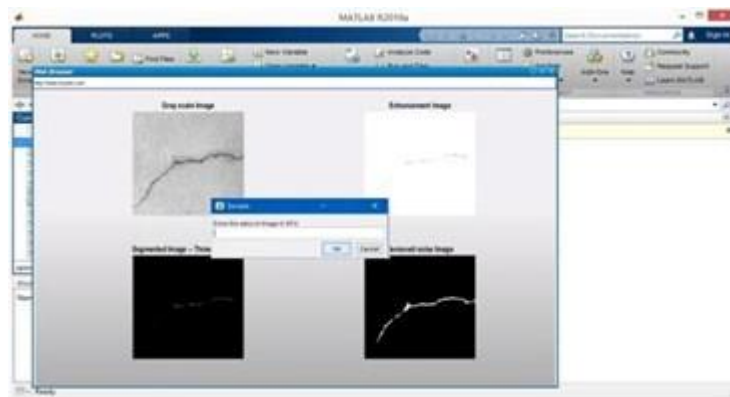




**Figure 18.** Procedure for Analysis of Performance of ACF detector



**Figure 19.** Processed image interface.



**Figure 20.** Prompt Command interface.



Figure 21. Result interface.

## PERFORMANCE EVALUATION

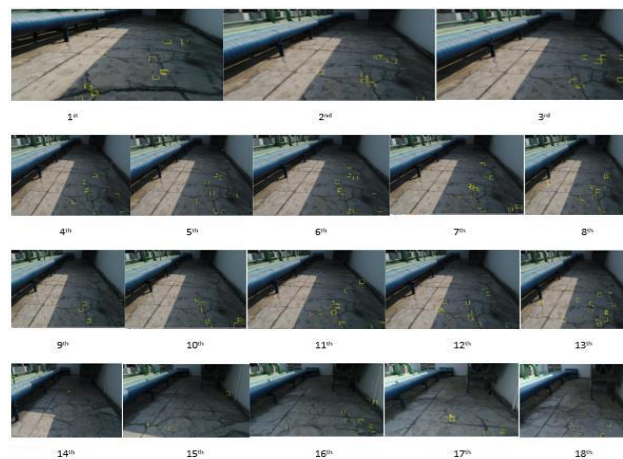
The comparison of the proposed work with the existing models for preciseness and crack severity using ML and UAV collected images has shown a great achievement in performance. The values of parameters obtained for the recognition of UAV collected images are remarkable. It can be observed that the highest recognition accuracy obtained by the proposed work is 85.94% by employing CNN based model in which automatic feature extraction takes place while others have achieved less accuracy. Our proposed model has achieved more accuracy in comparison with [12,13,14]. To evaluate the advantage of using Aggregate Channel Features (ACF) detector in the detection of crack on building, a collection of video data that used for training and real experiment is obtained by using UAV at Hospital Raja Permaisuri Bainun, Ipoh. A part of collected video data is abstracted for training purpose. In the training phase there are 4666 positive examples, 23330 negative examples and containing 2048 weak classifiers. The training phase undergoes 10 stages. The collected video data is used for testing purpose. This is because there is external disturbance such as video quality, light exposure, shadow, dirt on building, and crack look alike objects which will affects the learning of the neural network and the accuracy of detection.

### Evaluation of training efficiency

The training of neural network in real experiments implements the Aggregate Channel Features (ACF) detector with MATLAB's Ground Truth Labeler on a PC with Intel Core i5-5200U CPU and 8GB RAM. The training undergoes 10 stages with 5 'SamplingFactor' and 5 'NegativeSamplesFactor'. With 4666 positive examples, 23330 negative examples and containing 2048 weak classifiers, the training process takes about 58.62 mins to complete.

### Evaluating of detection performance

The experiments aim to shows the performance of Aggregate Channel Features (ACF) detector when detecting the crack on the video data collected. To test the performance of the crack detection in the testing data used, 100 selected images from the video data is abstracted from the real experiment to determine the accuracy of the system on crack detection. The 100 selected images are screenshots randomly from the real experiment. Figure 22 shows the selected 100 images from different angles.



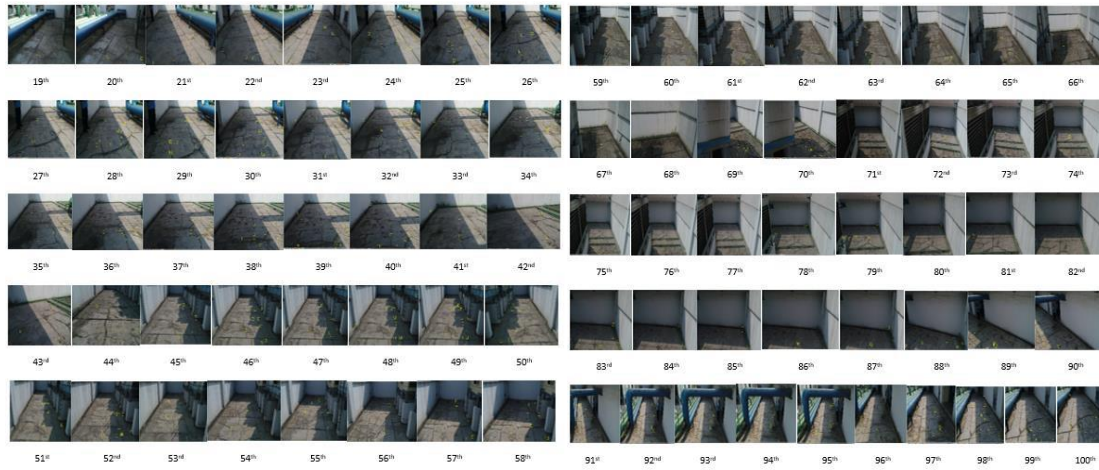


Figure 22. 100 selected images

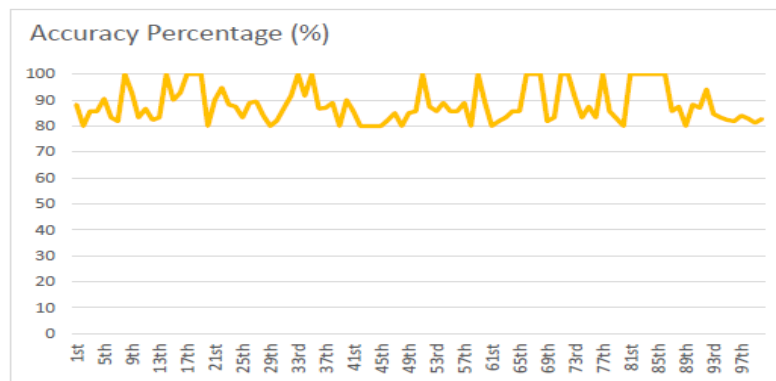


Figure 23. Accuracy Percentage of ACF detector

From the 100 selected images above shows that the average accuracy is in between 80% - 100%. The overall accuracy based on 100 selected images is 88.34% for crack detection by using Aggregate Channel Features (ACF) detector.

#### Training Efficiency:

We implement the method with Piotr's MATLAB toolbox on a PC with Intel Core i5-5200U CPU and 8GB RAM. With 4666 positive images and 23330 negative images in total, the training process takes about 78.62 mins for a detector containing 2048 weak classifiers. The time taken for training process is a bit longer due the hardware requirements.

#### Comparative Results:

When inspecting detections of the proposed crack detector, the testing result show that there are still a small amount of error in the detection of crack due to some external disturbance such as video quality, light exposure, dirt, and crack look alike objects. However, the result obtained show that the accuracy of crack detection is at average of 88.34%. Therefore, by the real experiment conducted had proved that Aggregate Channel Features (ACF) detector own the capability of fast feature extraction and strong performance ability.

## DISCUSSION

Crack detection in old buildings faces various technical challenges and difficulties, such as detection using physical inspection. UAVs can be used to collect images with precision, saving time and providing a higher quality image. Moreover, ML algorithms are widely utilised to provide further knowledge from the collected pictures about the severity of the damage created due to cracks. Hence, UAV and ML algorithms in congestion can be used to identify the cracks for best results. Even though many ML algorithms and techniques have been examined and evaluated for the same purpose, there is still a requirement to assess the quality of different machine learning methods and algorithms to determine whether the stated algorithm achieves the stated purpose target or not. This research work investigates and compares the current research in the fields of ML and UAV and hence motivates the development of enhanced and better algorithms with best performance. Furthermore, this work proposes a ML algorithm based on the compared assessments while utilising the UAV-based pictures of the old buildings.

In this research project, we proceed with the conventional neural network (CNN) implementation and keep the other machine learning algorithms (MLA), especially Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), CNN-SVM for comparison and their implementation as our future research for more accurate and best

results in our next stage of research. Therefore, here we use conventional neural networks (CNN) using MATLAB for our experimental implementation using the UAV.

## CONCLUSION

Previous works for ML for detection and classification of building cracks does not include the automated method of collecting the pictures using UAV. However, in our project implementation, we have utilized CNN model along with UAV utilization for the same purpose. Additionally, we select and compare the three deserving candidate techniques for crack detection, which use Machine Learning Algorithms and Unmanned Aerial Vehicles (UAV) and provide their performance evaluation for a better understanding, which motivates us for this research work. In this research project, an algorithm was developed by using MATLAB to analyse the crack detected effectively. An evaluation is conducted on the impact and the level of risk caused by the cracks in an automated manner. We evaluate the performance in depth and achieve an expected accuracy in case of crack detection using CNN-SVM from 82.94% to 85.94% for our use case of Hospital Raja Permaisuri Bainun, Malaysia. In our research project, we proceed with the conventional neural network (CNN) implementation and keep the other machine learning algorithms (MLA), especially Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), CNN-SVM for comparison and their implementation as our future research for more accurate and best results in our next stage of research. Therefore, here we use conventional neural networks (CNN) using MATLAB for our experimental implementation using the UAV.

## ACKNOWLEDGEMENT

The authors would like to thank UMP for funding this work under an internal grant RDU192622.

## REFERENCES

- [1] Wu, X., Jiang, Y., Masaya, K., Taniguchi, T., & Yamato, T. "Study on the Correlation of Vibration Properties and Crack Index in the Health Assessment of Tunnel Lining," *Shock and Vibration*, 2017.
- [2] Talab, A. M. A., Huang, Z., Xi, F., & HaiMing, L. "Detection crack in image using Otsu method and multiple filtering in image processing techniques," *Optik*, 127(3), 1030–1033. 2016
- [3] Adhikari, R. S., Moselhi, O., & Bagchi, A. "Image- based retrieval of concrete crack properties for bridge inspection," *Automation in Construction*, 39, 180–194. 2014
- [4] Hoang, N. D. "Detection of Surface Crack in Building Structures Using Image Processing Technique with an Improved Otsu Method for Image Thresholding," *Advances in Civil Engineering*, 2018. <https://doi.org/10.1155/2018/3924120>.
- [5] Smith, P., Reid, D. B., Environment, C., Palo, L., Alto, P., & Smith, P. L. Otsu\_1979\_otsu\_method. *IEEE Transactions on Systems, Man, and Cybernetics*, C(1), 62– 66. 1979
- [6] Mohan, A., & Poobal, S. "Crack detection using image processing: A critical review and analysis," *Alexandria Engineering Journal*, 57(2), 787–798. 2018
- [7] Kim, H., Ahn, E., Cho, S., Shin, M., & Sim, S.-H. "Comparative analysis of image binarization methods for crack identification in concrete structures," *Cement and Concrete Research*, 99, 53–61. 2017.
- [8] Thatoi, D., Guru, P., Jena, P. K., Choudhury, S., & Das, H. C. "Comparison of CFBP, FFBP, and RBF networks in the field of crack detection," *Modelling and Simulation in Engineering*, 2014. <https://doi.org/10.1155/2014/292175>.
- [9] Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., & Fieguth, P. "A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure," *Advanced Engineering Informatics*, 29(2), 196–210. 2015
- [10] Zakeri, H., Nejad, F. M., & Fahimifar, A. "Image Based Techniques for Crack Detection, Classification and Quantification in Asphalt Pavement: A Review," *Archives of Computational Methods in Engineering*, 24(4), 935–977. 2017
- [11] Rabah, M., Elhattab, A., & Fayad, A. (2013). "Automatic concrete cracks detection and mapping of terrestrial laser scan data," *NRIAG Journal of Astronomy and Geophysics*, 2(2), 250–255. 2013
- [12] Rau, Jiann-Yeou, K. W. Hsiao, J. P. Jhan, S. H. Wang, W. C. Fang, and J. L. Wang. "Bridge crack detection using multi-rotary UAV and object-base image analysis," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 42 (2017): 311.
- [13] Tedeschi, A., & Benedetto, F. "A real-time automatic pavement crack and pothole recognition system for mobile Android-based devices," *Advanced Engineering Informatics*, 32, 11–25. 2017. <https://doi.org/10.1016/J.AEI.2016.12.004>.
- [14] Chaiyasarn, K., Khan, W., Ali, L., Sharma, M., & Brackenbury, D. "Crack Detection in Masonry Structures using Convolutional Neural Networks and Support Vector Machines," Proceedings of the 35th ISARC, Berlin, Germany. pp. 118-125. 2018.
- [15] Santos, R., Ribeiro, D., Lopes, P., Cabral, R., & Calçada, R. "Detection of exposed steel rebars based on deep-learning techniques and unmanned aerial vehicles," *Automation in Construction*, 139, 104324. 2022.
- [16] Kung, R. Y., Pan, N. H., Wang, C. C., & Lee, P. C. "Application of deep learning and unmanned aerial vehicle on building maintenance," *Advances in Civil Engineering*, 2021.
- [17] Ko, P., Prieto, S. A., & de Soto, B. G. "ABECIS: An automated building exterior crack inspection system using UAVs, open-source deep learning and photogrammetry," In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 38, pp. 637-644). 2021. IAARC Publications.
- [18] Saeed, M. S. "Unmanned Aerial Vehicle for Automatic Detection of Concrete Crack using Deep Learning," In 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST) (pp. 624-628). IEEE.
- [19] Silva, W. R. L. D., & Lucena, D. S. D. "Concrete cracks detection based on deep learning image classification," 18th International Conference on Experimental Mechanics (ICEM18), Vol. 2, No. 8, p. 489. 2018. MDPI AG.
- [20] Bhowmick, S., Nagarajaiah, S., & Veeraraghavan, A. "Vision and deep learning-based algorithms to detect and quantify cracks on concrete surfaces from UAV videos," *Sensors*, 20(21). 2020.



- [21] Best-Rowden, L., Han, H., Otto, C., Klare, B. F., & Jain, A. K. "Unconstrained face recognition: Identifying a person of interest from a media collection," *IEEE Transactions on Information Forensics and Security*, 9(12), 2144–2157. 2014 <https://doi.org/10.1109/TIFS.2014.2359577>
- [22] Dollár, P., Tu, Z., Perona, P., & Belongie, S. "Integral channel features," *British Machine Vision Conference, BMVC 2009 - Proceedings*, January 2009. <https://doi.org/10.5244/C.23.91>
- [23] Bourdev, L., & Brandt, J. "Robust object detection via soft cascade," *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, II(July 2005)*, 236–243. <https://doi.org/10.1109/CVPR.2005.310>