Coventry University

DOCTOR OF PHILOSOPHY

Deep learning approach for road pothole detection using accelerometer and imagery data

Pandey, Anup Kumar

Award date:
2021

Awarding institution:
Coventry University

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of this thesis for personal non-commercial research or study
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission from the copyright holder(s)
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 06. Jul. 2025

Deep Learning Approach for Road Pothole Detection Using Accelerometer and Imagery data

By

Anup Kumar Pandey

BSc. (Hons), MSc (Distinction), MBA

A thesis submitted in partial fulfilment of the University's requirement for the degree of Doctor of Philosophy

Institute for Future Transport and Cities

Faculty of Engineering, Environment and Computing

Coventry University.

April 2021

ETHICAL APPROVAL CERTIFICATION

Certificate of Ethical Approval

Applicant:	Anup Pandey
Project Title:	Automatic Detection Of Road Defects To Inform Predictive Maintenance Actions

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Low Risk

Date of approval:	05 Apr 2021
Project Reference Number:	P121393

DECLARATION

Content removed on data protection grounds

ACKNOWLEDGEMENTS

I have been extremely fortunate to have Dr Rahat Iqbal as my Director of Study. Dr Iqbal has been encouraging and supportive throughout my PhD. I still remember it clearly when I had approached Dr Iqbal for the very first time to discuss my research project. He was very welcoming and listened to me carefully to understand my aspirations. He allowed me to work on this project. Dr Iqbal is always positive and inspires his students to achieve higher. He has not only advised me in a professional capacity but has also helped me facilitate uninterrupted work on this project. He has consistently welcomed my thoughts and guided me positively. This research project could not have been completed in time without his positive and encouraging attitude. I am grateful for his support. Given a chance, I will always prefer to work with Dr Iqbal. Thank you, Dr Iqbal.

Dr Palade is another gem I met at Coventry University. Dr Palade taught me the basic concept of Deep Learning and have always guided me to achieve a higher technical competency. Dr Palade has always maintained his calm composure and answered my not so intelligent questions very effectively and patiently. Dr Palade has been a great help to complete this research project. Thank you, Dr Palade.

I would like to thank my industrial supervisor, Dr Tomasz Maniak, who has helped me understand the complex technology-related to this project. I was very impressed with Dr Maniak's technical competency from our very first meeting. He has always been there whenever I needed any technical help. I thank Dr Karyotis for his valuable comments on my thesis and help to make it better.

I would also like to thank Interactive Coventry Ltd for sponsoring my PhD and to all the unknown people whose works I have referred for this project.

DEDICATION

to my family and friends who keep me positive and inspired

PUBLICATIONS AND CONFERENCE

1. Kumar A., Iqbal, R., Maniak, T., Karyotis, C., Palade, V., (2021) : "Convolution Neural Networks for Pothole Detection of Critical Road Infrastructure", Journal of Computers and Electrical Engineering, Elsevier (in press).

2. Kumar A., Iqbal, R., Maniak, T., Karyotis, C., Palade, V., (2021) : "Deep Neural Networks Based Approach for Road Damage Detection Using Sensor and Image Data", Journal of IEEE Internet of Things, IEEE (in press).

ABSTRACT

Highways are critical infrastructures with a significant impact on economic and social prosperity. The UK's highways carry over 65% of domestic freight movement and 90% motorised passenger travel. For congestion-free travel and road users' safety, effective identification of potholes and maintaining the roads in good condition is crucial. Road potholes on the road cause inconvenience to commuters, delay in delivering product and services leading to a loss in the national GDP. Potholes on the surface of the road can lead to physical injuries and be a cause of death. Highway maintenance is essential; however, it has become challenging to keep highways in good condition due to increasing traffic, insufficient budget, and lack of human resources. Effective detection of potholes and timely maintenance of roads is crucial for road users' health and safety. The current pothole detection methods require manual inspection of roads, performed with custom sensors installed on specially adapted vehicles. The procedure is time-consuming and labour extensive. The current pothole methods are inefficient and lagging to keep pace with the demand to keep roads in good condition. Few methods use Machine Learning models with sensory and imagery data separately to classify roads. However, the Machine Learning-based sensor data model fails to differentiate between road anomalies and hinges. The Machine Learning model with imagery data has a low defect rate when the road is full of water. The model fails to differentiate between real road anomalies and thin dark objects, similar to a road anomaly. Furthermore, in order to address the delay in road surface information sharing, the Internet of things (IoT) can be used. IoT is an emerging technology and has the potential to provide an efficient and cost-effective solution to road pothole detection. In this thesis, training and testing data were collected using a smartphone as well as downloaded from the Internet (google search) to imitate crowd data sourcing.

To address the issues of a sensory data-based model and imagery data-based model this thesis proposes a novel fusion model based on Convolution Neural Networks (CNN). The fusion model will take sensory and imagery data as two inputs and predict an output considering both sensory and imagery data. This study proposes a cloud-based crowd data sourcing method to collect data. In the cloud based crowd data sourcing method, the road users from across the world will be able to upload images of road anomalies on the dedicated cloud server. The data from the server will be downloaded at the backend to process and detect road anomalies. The proposed method will tag potholes with geographical location and send a notification to the

road users who have opted for it. In this study, the images were collected using an iOS smartphone as a dashboard camera while accelerometer data was collected through a dedicated app on an iOS smartphone. The fusion model has achieved 87.20% precision, 92.70% recall and 89.9% F1-Score. The results show that utilising a fusion Convolution Neural Networks modelling approach with mixed input, image, and accelerometer data can produce better results. The proposed method is simple, cost-effective and computationally less extensive.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Potholes are not a new issue. A reader complained to the New York Times in 1910 that "a steady succession of the pothole was rendering travel a burden rather than a joy" (BBCNews, 2019). Potholes endanger road users and may cause significant harm not just to vehicles but also to the drivers. The expense of repairing potholes is high and needs special budgeting. According to the House of Commons report, there is a lag of nearly 14 years to complete the essential repair works (Report, 2016). The report mentions that the average number of potholes filled every year in England is nearly 16,000, and in London alone, the number is 4,099. The cost to fill these potholes is about £110 million and £11 million, respectively. According to the United Kingdom's Department of Transport, bad road conditions were responsible for almost 12% of all road accidents in 2016 (Report, 2016). A survey conducted by AA suggested that many drivers had their car damaged due to potholes, and the cost to fix these damages was nearly £684 million over a period of just 12 months (Report, 2013). Traffic jams caused the UK economy £8 billion in 2016 (RAC, 2016).

The road users want a pothole-free and smooth road to commute. Maintenance of roads and highways is vital for effective traffic management. Highway and road maintenance is essential for adequate traffic flow and the economy in general. Maintaining such a vast network of roads necessitates both expertise and funds. The UK government has set out a long-term funding program to create smooth, smart and sustainable roads, in order to ensure that roads and highways are properly managed and maintained. However, there is an inherent challenge in this task since the UK has a large network of roads of 262,300 miles (422,100 km).

Over the years, specialists are used to survey roads and the accountable authorities repair these roads. These procedures are labour-intensive, expensive, and time-consuming, and they have significantly increased the workload of these professionals (Jahangiri, et al., 2015). Also, this method is not able to meet the increasing demand to keep roads in good condition. As mentioned above, there is a severe delay to repair road damages. The authorities blame shortage

of human resources and funding for the delay in repairing these potholes. Furthermore, experts' inspections may not be consistent to their subjective visual perceptions.
With time, the road network is growing, and the number of road users too. As technologies are improving, more and more systems are becoming autonomous and are adopting Artificial Intelligence to deliver better services, thus attempting to make consumers' experience more enjoyable and safer. There are many well established companies and start-ups working on driverless automobiles. The number of self-driving vehicles on the road is growing faster than expected, and their market will be worth £ 41.7 billion by 2030 (Gov-Report, 2021). Such automobiles will require robust systems to identify the road conditions and texture to deliver better services. To keep pace with the demand, it is inevitable to start using the latest technology to keep the road well maintained and safe. The early-stage detection of road surface anomalies will facilitate effective maintenance.

The latest technological developments are enabling more systems to support pothole detection and provide autonomy, by utilising Deep Learning models. Several studies have proposed the use of artificial neural networks to inspect and identify damage to the road surface (Sabanovic & Zuraulis, 2020) (Basavaraju, et al., 2020) (Huidrom & Das, 2013). A cost-effective Machine Learning method can be used to detect road anomalies (Basavaraju, et al., 2020).

The accuracy of any Deep Learning model depends on the quality of the dataset which is used to train the model. In the past, data collection has been a challenging task requiring a dedicated recording device to collect the data in order to develop Machine Learning models. However, with the development of technologies, smartphones are ubiquitous and can record good imagery data, sensory data, and GPS details. Furthermore, the progress in cloud technologies is making the job of data collection, storage and processing much easier. (Burke & Srivastava, 2006) discussed participatory crowd-based data collection for sensing, an emerging methodology using modern smartphones that are widespread and have great sensing features. (Zang & Jie , 2018) used sensor data collected by smartphones to map road surface roughness based on the international roughness index. Machine Learning methods are proven to be viable and cost-effective for road surface anomalies detection. (Artis, et al., 2011). Hence, it is imperative to automate road surface inspection using state of the art Machine Learning techniques. This study investigates how smartphone sensors may be used to detect potholes.

To address the road pothole issue, this Thesis proposes a cloud-based architecture that uses a fusion model based on Convolution Neural Networks (CNN). The fusion model will utilize imagery and accelerometer data from smartphones as inputs to identify road potholes. The proposed method uses the crowd data collection method to collect imagery and sensor data along with their geographical locations. The road users will be able to use their smartphones to collect imagery and accelerometer data with their geographical locations, time stamp details and upload data to a cloud-based server. At the back end, the server will use a trained CNN model to identify road potholes and notify the users who have subscribed to this dedicated service. The dataset of road potholes will be continuously updated with new road potholes in real-time. Public authorities can use the pothole database for keeping the road in good condition. Autonomous vehicles or companies providing route planning services can also benefit from the approach and dataset developed by this study. The proposed method does not require a specialised computer vision device or high computational capability.

The rest of the chapter is organised as follows. Section 1.2 discusses research motivation. Section 1.3 discusses research question, Section 1.4 discusses research aim and objectives. Section 1.5 sets research scope. Section 1.6 presents research contribution. Section 1.7 describes research approach and its various components. Finally, Section 1.8 presents the layout of this Thesis.

1.2 Research Motivation

The thrust to develop an in-depth understanding of Deep Learning models and its real-world applications that are beneficial to society has motivated the undertaken research. Towards, this endeavour, various approaches relating to Deep / Machine Learning were investigated and various real-world applications were explored. However, a well acknowledged outstanding issue of pothole detection was identified for the undertaken research. This research is aimed to develop a novel Deep Learning Model that can automatise road surface monitoring and address the road pothole issue. The initial research was carried out to understand which methods are currently used to monitor the operational state of roads with respect to road damage detection. While conducting the research it was noted that road surface anomaly detection methods have been studied for many years. However, the methods used to detect road surface anomalies and potholes have changed over time. Traditionally, there are three approaches to monitor road surface:

3D reconstruction, vibration and computer vision-based (Buza & Omanovic, 2013). The 3D reconstruction requires a 3D laser scanner, which scans the surface and makes an accurate model compared with the base model to detect anomalies (Chang, et al., 2005). However, such laser scanners are very costly, and the methods are focused on the local accuracy of the 2D scan (Kim & Ryu, 2014). UdedaK (1995) and ItakuraY (1982) proposed the polarization method to calculate the difference between horizontal and vertical polarisation. However, the polarisation filters may affect the quality of the images, hence reducing the detection accuracy (UedaK, 1995), (ItakuraY, 1982). Gailius (2007) proposed variation in ultrasonic noise method, due to friction between tire and road surface, method to detect black ice on the roads can be found using tire to road friction ultrasonic noise algorithms (D. Gailius, 2007). The experiment was started to assume that specific differences in the acoustic noises made when tire hits, interact and pull off the road surface. The spectrum from 50 to 100 kHz was chosen for the experiment. The study could not establish any correlation between noise and the presence of black ice on the road. However, it suggested other atmospheric factors such as temperature and humidity, which could be used and potentially studied. As discussed above, these methods and technologies are expensive and resource intensive and therefore the motivation is to develop a cost effective and reliable solution for pothole detection using state of the art Deep Learning Approaches.

1.3 Research Questions

This research aims to address the following research questions:

1. Is it possible to develop an automated road surface detection application?
2. Can we use combinations of different data (image and sensory data) to develop more effective pothole detection applications?
3. Is Deep Learning able to exploit diverse data sources and achieve state of the art (SOA) performance in pothole detection?
4. Can this research provide a simple and cost-effective method for pothole detection?
5. Is it possible to utilize standard smartphones to collect meaningful combinations of data that can be used effectively for pothole detection?

1.4 Research Aim

The main aim of this research is to investigate novel methods for pothole detection using Machine Learning approaches. To answer the above research questions (section 1.2), the following objectives are set to be achieved through the conduct of this research:

1 To explore the issue of pothole detection and highlight its significance for health, safety and the economy.
2 To investigate the state-of-the-art Machine Learning approaches for pothole detection.
3 To identify various existing data sets related to pothole detection.
4 To capture sensory and imagery data of adequate volume and quality for pothole detection.
5 To apply custom Deep Learning based approaches to identify and detect potholes automatically.
6 To objectively validate the performance of the proposed approach.

1.5 Research Scope

The research scope is limited to design a Deep Learning model to address the real-world issue of pothole detection. The research scope was limited to using hardware such as a simple smartphone, and a publicly available software application to collect data for the project in order to demonstrate the usefulness of everyday technology for pothole detection. Two streams of data namely sensory and imagery data were collected using smart phone and the dedicated software. This research project did not cover the development of any hardware used in the project. However, to achieve the research aim, a comprehensive literature review in the area of Deep Learning, Computer vision and application of sensory data in the real world was conducted. The thesis also discusses an architecture which can be used to report road potholes and share information with the stakeholders.

1.6 Research Contribution

This research has developed a novel fusion model based on Deep Learning algorithms to process sensory and imagery data for pothole classification and detection. This is one of the first comprehensive approach which has been developed using state of the art Machine Learning algorithms for data fusion. This thesis has made the following contributions:

1. Development of an optimal computer vision model using two-dimension Convolution Neural Networks to detect road pothole.
2. Development of an optimal Deep Leaning model of one dimension Convolution Neural Networks to detect pothole using sensory data (accelerometer data).
3. Development of a fusion Deep Leaning model based on Convolution Neural Networks to detect pothole using imagery and sensory data

In order to achieve the above, the thesis proposes following architectures to collect data and share road pothole information.

4. Development of a cloud-based architecture to report road potholes and share of information on road potholes with various stakeholder.
5. Development of a cloud-based architecture to collect data using the crowd sourcing method.
6. Design of a methodology to collect sensory data (Accelerometer and GPS) using a generic application available on an iOS smartphone and collect imagery data using a smartphone camera.

1.7 Research Approach

The proposed research is carried out by following a systematic methodology as shown in Figure 1. The research started with a vague idea to develop a Deep Learning method to detect road potholes which are a big nuisance to everyone. The first stage was to understand the scope of the issue and available technologies, methods and techniques. Also, an investigation was conducted to review the available commercial tools which can help the stakeholders to address the issue of pothole detection. Along with exploring the commercial space for a solution, the literature review was conducted to investigate the state-of-the-art methods which can help to automate road surface monitoring and road pothole detection. The first stage covered extensive research and literature review. By the end of the first stage, there was a clarity on how the research can contribute to resolve the pothole detection issue more effectively. It was decided to use Deep Learning Method to develop a state-of-the-art system which can effectively and efficiently detect road potholes.

In the second stage, as shown in Figure 1, Imagery data was collected and pre-processed. The various image processing and augmentation methods were used to prepare the imagery dataset as described in CHAPTER 4. The prepared imagery data was used to run experiments with various two-dimensional Convolution Neural Networks (2D-CNN) models. These models have varying hidden layers and use images of varying sizes as discussed in section 5.2. In the second stage, a one-dimensional Convolution Neural Networks (1D-CNN) was developed to process sensory data in order to detect road potholes. In this experiment, varying number of hidden layers, different dropout and different kernel size were used to determine an optimal model. The optimal model was decided based on training and testing accuracies, namely Average Precision Rate and Average Recall Rate. In the second stage of the research the experimental results were documented. By end of the second stage of the research, many satisfactory experiments with Deep Learning Models of 2D-CNN and 1D-CNN were completed. During this time, the review of the literature continued to keep up to date with the latest developments in the related field and the latest publications. Weaknesses of imagery method and sensory method were investigated as discussed in CHAPTER 2. In order to address the limitation a fusion model was developed which can address the weaknesses of computer vision model and sensory data model.

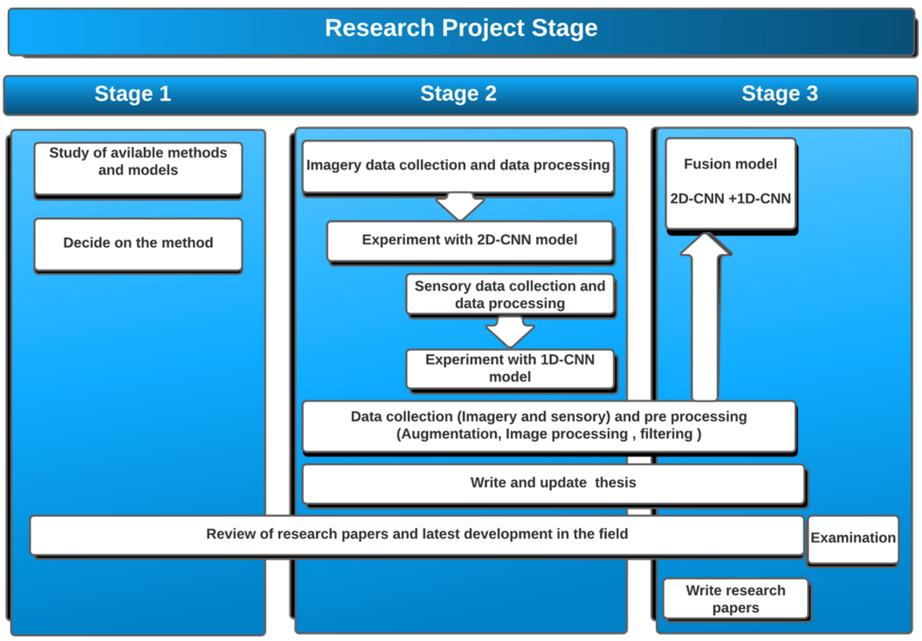

Figure 1:Schematic diagram of research model development

In the third and the last stage of the research project, a fusion model was developed, as described in Section 3.1, based on the results from the stage 2. The fusion model was trained based on imagery and sensory data for pothole detection as described in section 5.4. This has led to the completion of this thesis and research articles for research results dissemination.

1.8 Thesis Outline

The rest of the thesis is organised as follows.

CHAPTER 2

Chapter 2 presents a literature review on the state-of-the-art approaches relating to road pothole detection. The literature review started by reviewing the general method employed to detect road surface anomalies before computer vision, and Machine Leaning arrived in the field. Section 2.4 reviews literature related to the use of imagery data with Machine Learning. Section 2.5 covers related work associated with the use of sensory data collected with the help of smartphones to be used in Machine Learning methods. This section covers relevant work which have used the fusion method to use imagery and sensory data together.

This chapter also discusses the theoretical part of the research and briefly explains Machine Learning, Neural Network and the architecture of Convolution Neural Networks (CNN).

CHAPTER 3

Chapter 3 discusses the research methodology followed. This chapter discusses the proposed cloud-based architecture, how a road user captures a road pothole using a smartphone, how the potholes are uploaded on a cloud-based server at the backend, how data is processed to identify a road pothole and eventually how the road pothole data base is updated in real-time. The detailed research methodology of three experiments has been described in this chapter. These experiments include experiment 1 for imagery data; experiment 2 for sensory data and experiment 3 for mixed imagery and sensory data.

CHAPTER 4

Chapter 4 discusses the methodology of imagery and sensory data collection. It discusses smartphone placement on the windshield, smartphone orientation on pothole detection, the impact of speed and motion blur on pothole detection. Later it discusses about how the imagery data set was prepared after investigating the raw images, removing unwarranted images, resizing the images before labelling the images with damager classes. Once the imagery data was labelled, the image augmentation technique was applied to the images to increase dataset size. A similar process was used on the sensory data set. The sensory data samples were resampled to make sure that they all are of the same length. Later, a Python script was used to match sensor data's time and GPS location with imagery data to label sensory data into damage classes. After labelling was completed, sensory data were subject to data augmentation technique such as variation in amplitude and permutation to increase the sensory dataset size. The end data set was split into 70% for training, 15% for validation and 15% for testing.

CHAPTER 5

Chapter 5 discusses the three experiments, shown in Figure 16 Experiment-1 uses two-dimensional Convolution Neural Networks and imagery data to detect road pothole. Several models with varying combinations of classes and hidden layers have been tried in this experiment. The output graph and accuracy have been discussed in detail. Experiment-2 uses a 1D-CNN model with sensory data to detect road potholes. Many 1D-CNN models with different hidden layers, kernel size and dropout have been tested. The results have been discussed and illustrated with the help of graphs. The third and the last experiment-3 explores the design of a fusion model using 2D-CNN and 1D-CNN. The fusion models take imagery data and sensory data as inputs and produces detection result. The results of training and testing the fusion model have been discussed in detail in this chapter.

CHAPTER 6

Chapter presents conclusions and future work. This chapter discusses the main contribution of the Thesis and its limitations.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This research is inspired by the acknowledged success of Deep Learning (DL), computer vision and big data analytics in several demanding application domains, such as medicine and fault detection. This thesis focuses on pothole detection and reviews the relevant state of the art approaches and techniques concerning Deep Learning. This chapter will investigate how various Deep Learning approaches can be combined to develop a hybrid Deep Learning method in order to overcome the weaknesses of individual approaches. An insight will be taken from this chapter to address the computational and modelling challenges. Both Machine and Deep Leaning methods will be reviewed in this chapter. Additionally, the basics of Machine Learning and Convolution Neural Networks will be outlined based on the existing literature. Most importantly, this chapter will review the literature related to Machine Learning approaches concerning imagery and sensory data.

The rest of the chapter is organised as follows. Section 2.2 discusses Machine Learning approaches and their types, namely supervised learning and unsupervised learning. Various applications of Machine Learnings will also be reviewed. Section 2.3 will discuss Deep Learning approaches. The basic architecture of Neural Networks and Convolution Neural Networks (CNN) are outlined in this chapter in order to understand the underlying concepts, its usage and benefits. Section 2.4 will be focussed on Machine Learning approaches for image analysis. Section 2.5 will cover Machine Learning approaches and their applications in the areas of sensory data analysis. At the end of this chapter, in section 2.7chapter summary will be presented.

2.2 Machine Learning

Machine Learning is a field of Artificial Intelligence and has become a core component in digitalisation solutions that have gained considerable interest in the modern arena. It is the most effective data analytics method for forecasting by creating models and algorithms. It also helps in the detection of latent patterns or data characteristics centred on prior learnings and trends. Machine Learning is widely used in pattern recognition, natural language processing, traffic congestion prediction, computer games, share market prediction, E-mail spam filtering and many more. These issues can be divided into classification, clustering and regression. Based on the kinds and categories of training data accessible, one may need to choose between unsupervised learning methods and supervised learning to implement the required Machine Learning algorithm. A high-level overview of Machine Learning methods is given in Figure 2.

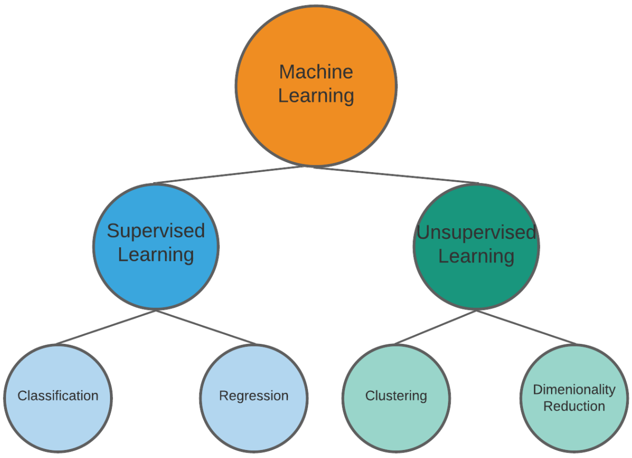

Figure 2: Machine learning

In the case of unsupervised learning, the model is not provided with labelled data during the training. The model tries to find the hidden pattern using statistical properties of the data set and cluster the dataset into different categories. Fuet al (2015) proposed an unsupervised learning method for high spatial resolution remote sensing images scene feature extraction and feature learning. Their method first extracted feature then completed feature learning before using the Support Vector Machine (SVM) to classifying the dataset. (Fu, et al., 2015). They provided raw data and used the RGB pixel intensities as features and used K-means clustering to cluster the images. Figure 3 shows the process to extract feature and the use of SVM to classify (Fu, et al., 2015).

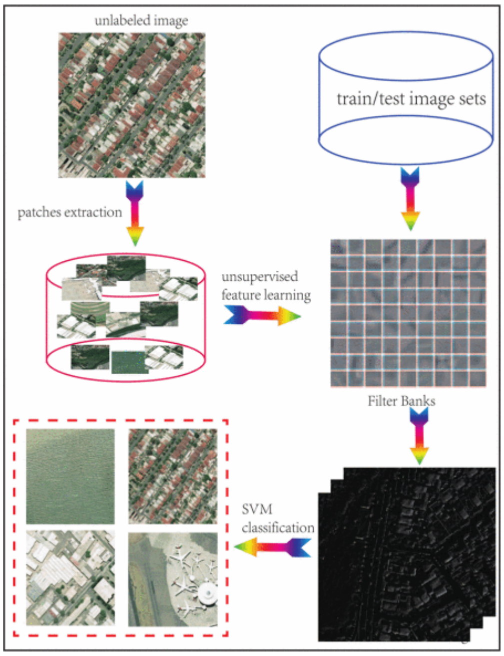

Figure 3:Unsupervised learning process

In the case of supervised Machine Learning, data set is labelled and include input and desired results. The data set contain labelled input and corresponding result for the input. Supervised learning has developed itself as a significant class label distributor of predictor features. Supervised learning is faster and accurate as the dataset is labelled. Three commonly used supervised ML algorithms are Support Vector Machine (SVM), Decision Tree (DT) and Neural Networks.

SVM was suggested by Vatnik in 1992 to be used in non-linear data. The development of a hyperplane for classification is fundamental to the SVM definition. According to the principle of SVM, if training data is divided into n categories, then the SVM training algorithm will find a super plane to classify these n classes. Raj, et al. (2019) discussed the use of the SVM approach to classifying underwater images. They discussed how the accuracy of underwater photographs is affected by a variety of physical processes such as backward and forward scattering and light absorption. They also discussed a variety of other considerations related to the complexity of underwater picture classification. First, the density of the water is directly proportional to the object's uncertainty, i.e., as the depth rises, so does the object's uncertainty. This confusion stems from the reality that certain marine animals have the ability to blend in with their surroundings underwater, resulting in subtle shifts in the image's context. Therefore, seeking a successful mix of feature extractor and classifier for a dataset with incomplete depth knowledge becomes a difficult challenge. Second, the expense of underwater equipment is a significant consideration. Third, recognizing and selecting good features from the dataset is a challenging task. Regardless, it is a prerequisite for object classification. Later they used SVM to classify the images into 7 categories. The proposed approach achieves an accuracy of 93% (Raj & Murugan, 2019)

A Decision Tree is a supervised training algorithm and can be used for classification as well as regression. For this reason, sometimes the Decision Tree approach is also called CART. Decision Tree uses a mathematical algorithm to generate a Decision Tree using training data. It then uses the generated tree for classification (Shen, et al., 2011). Figure 4 shows detail of training and classification using a decision tree.

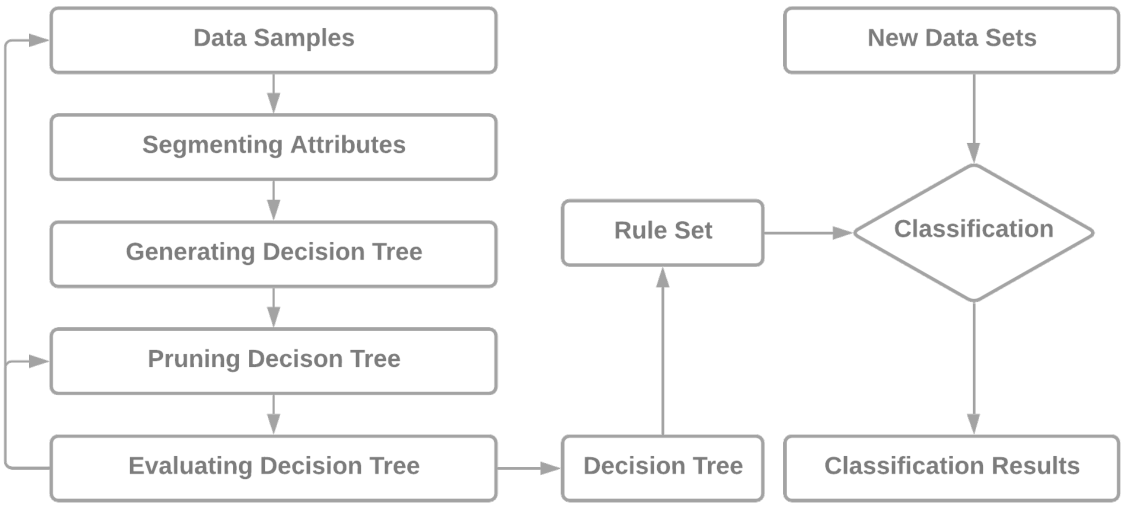

Figure 4:Decision tree process

Machine Learning is not limited to classification but has also been used to solve regression, clustering, dimensionality reduction and many more issues. However, in this thesis the use of Machine Learning in the Image classification issue has been discussed. Image classification is a difficult challenge, but it can be made easier with the supervised Machine Learning approach. In general, the classification of images into separate groups requires two steps: the first is the extraction and recognition of features, followed by the classification of images depending on the obtained features. Machine Learning is one of the techniques for accomplishing this. The next section will discuss Deep Learning which is a subclass of Machine Learning and has emerged as dominant approach in image classification field.

2.3 Deep Learning

There have been many studies proposing the use of the Deep Leaning (DL) method to identify road pothole. The DL technique is gaining popularity among researchers and has contributed to the solution of many issues. DL has been used to develop state of the art solutions for many real-world issues. Deep Learning has an excellent track record in tasks, such as image recognition, data analytics for a particle accelerator and speech recognition. DL has achieved excellent performance in forecasting and classification issues compared to other traditional methods (Karyotis, et al., 2019). The Deep Learning method can be put into two categories based on their training method. The first one has supervised learning in which labelled data is used and the second is unsupervised learning in which Deep Learning model extract hidden pattern and no labelled data is used to train the model.

Deep Learning networks are made of Neural Networks with many hidden layers. Artificial Neural Network (ANN) is a nonlinear statistical model tool that can found a complex pattern between input and output variables. ANN is based on a biological neural network. ANN is developed based on the understanding, structure and functions of human brain.

ANN is comprised of a minimum of three layers: Input Layer, Hidden Layer and Output layer. The number of hidden layers could be increased based on the complexity of the issue. Each hidden layer will have many nodes, and they will be connected with the node on the previous or to the next layer. The number of nodes in the output layer would be based on the number of classes the model is trying to identify. A fundamental Artificial Neural network is as shown in Figure 5.

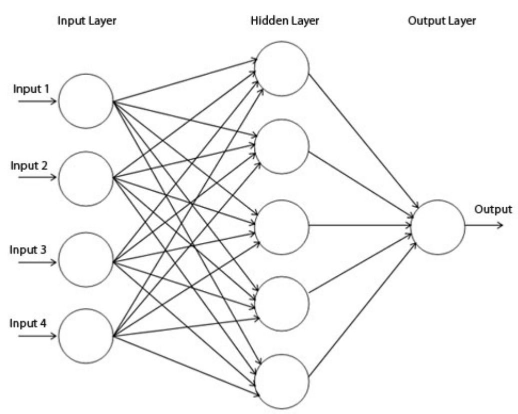

Figure 5:Basic artificial neural network

Convolutional Neural Network is one of the most famous Deep Learning method. Convolution Neural Networks in combination of Image processing could be used to find patterns in Images and classify them.

Convolution Neural Network is a type of Neural networks inspired by animal' visual cortex. (Yamashita, et al., 2018). (Fukushima, 1980) introduced Noncognition which is similar to convolutional Neural Networks. (Lecun, et al., 1998) discussed the issues of multi-layer neural networks in object recognitions. The large image size required a large dataset to learn the distinctive features wherever they appear on the input. The second issue with a fully connected architecture was that it ignored the topology of the information. However, Convolution Neural Networks extracted local features by restricting the receptive fields to be local. Convolution Neural Networks were predominately designed for image classification.

Convolution Neural Networks learn features from two-dimension images and produce an output. The Convolution Neural Networks architecture has three main layers: Convolution Layer, Pooling Layer and Fully Connected Layer. The adjacent two layers of neurons are connected, forming a directed acyclic graph. Convolution Neural Networks reduces the complexity of computation in the training of the network by sharing the weights. The model's capacity and complexity can be changed by changing the number of Convolution Neural Networks layers and their organization. The Convolution Neural Networks reduce the learning complexity of the model (Ahmed & Tao, 2017).

The architecture of Convolution Neural Networks has been shown in Figure 6. A basic Convolution Neural Networks has Convolutional Layer, Maxpooling layer, dropout layer, fattening layer and dense layer as shown in *Figure 6*. All these layers perform specific works, which has been discussed in the details in the next section.

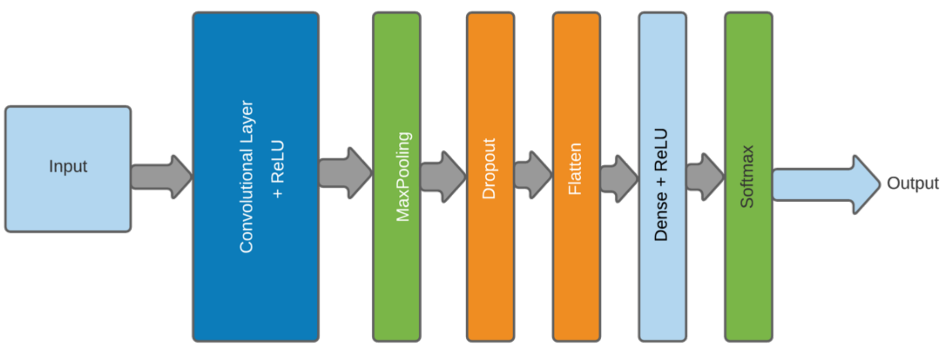

Figure *6* : Convolution neural networks architecture

The convolution layer is the fundamental layer of the architecture of Convolution Neural Networks. The convolution layer is responsible for feature extraction by a convolution operation. The convolution layer uses learnable filters called kernel to perform feature extraction. The kernel is defined as WxHxChannels.; where W and H are the width and height of the channel, and Channels shows the RGB (colours) of image input. The kernel is then slid (convolve) across the input image's height and width to calculate element-wise product between the input and entries of the filters at any position and summed to get output value for the position called a feature map. This process is repeated to form several feature maps. The kernel size could be different according to the application at different convolution layer *Figure 7*, *Figure 8* and *Figure 9* show how the convolution process is performed with a kernel size 3x3. The images have been taken from (Yamashita & Nisho, 2018).

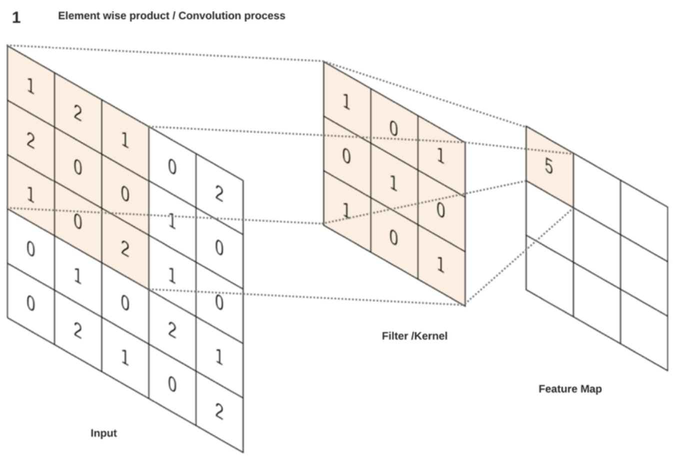

Figure *7*: Convolution process with a 3x3 kernel

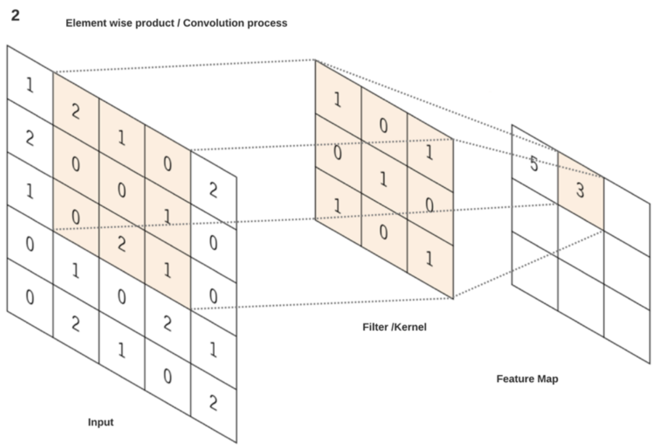

Figure *8*:Convolution process with a 3x3 kernel

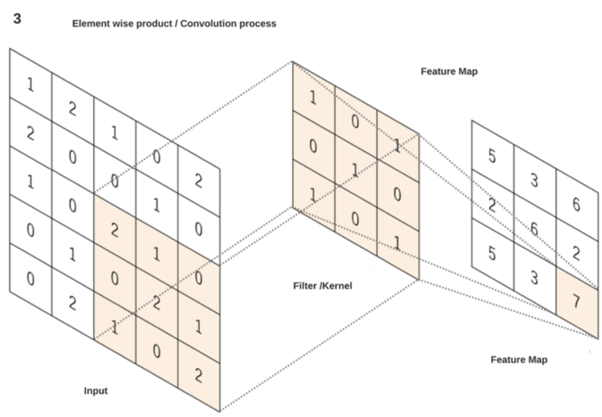

Figure *9*:Convolution process with a 3x3 kernel

The convolution operation described above does not allow the kernel's centre to overlap the end of input. To resolve this issue, padding is applied around the input. Mostly zero paddings are used around the input, so the kernel's centre can overlap with the end of inputs. The padding process can be noticed in *Figure 11*.

The activation function is used to activate a node. Activation functions transform inputs to output based on the function which has been used during the activation. There are many non-linear activation functions such as Sigmoid, Tanh, ReLU and SoftMax. SoftMax: It is a generalised form of the sigmoid and used in multi-class classification. It produces a value between 0 and 1 and so it is used as the final layer in classification models. ReLU: Rectified Linear Unit is defined as $F(y) = max(0,y)$. It is the most commonly used activation function. It handles the issue of vanishing gradients. This layer makes any negative number zero.

The dense layer is deeply connected and each neuron in the dense layer receives input from all neurons in the previous layer. The dense layer takes input from the previous layer and performs a matrix-vector multiplication and uses activation function and bias to produce output fed to the next layer in the architecture.

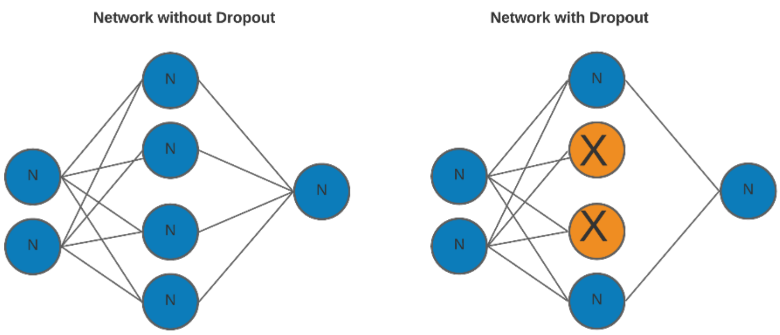

Figure *10*: Dropout layer

Dropout functionality was first proposed in 2012 (Hinton, et al., 2012). The dropout layer performs the dropout, which is a method to drop selected neurons during training randomly. It is used to avoid overfitting during the training of neural networks. The test error can be minimized using many different networks, but this is computationally extensive and expensive. The dropout allows training a large number of various networks in a reasonable time. (Hinton, et al., 2012).The layer Max pooling is a down sampling process in Convolution Neural Networks. Max pooling picks the maximum value within a matrix. Max pooling is used to reduce the input size and hence several parameters to lower the computation.

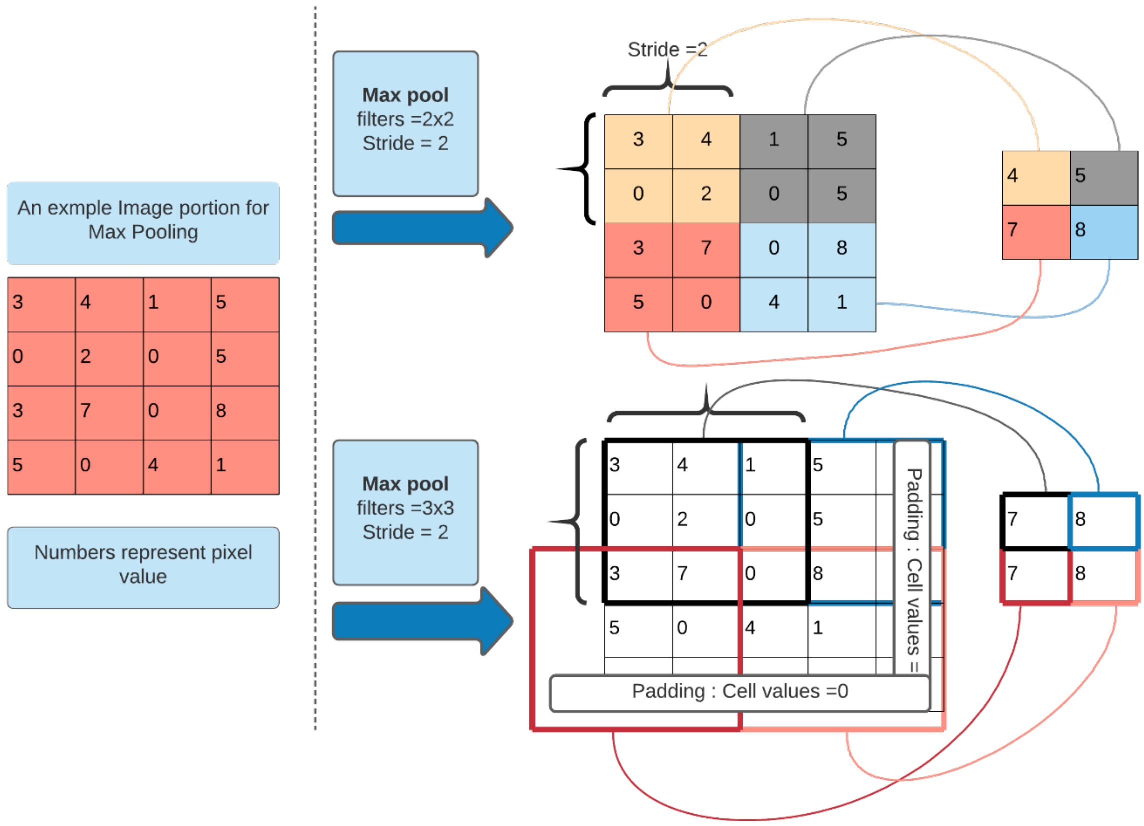

Figure *11*: Max pooling layer

Flatten: Fattening is used to convert data into a one-dimensional vector to feed into the next layer. Batch Normalisation is a layer in Convolution Neural Networks which normalise the output of the previous layer. It was proposed by (Ioffe & Szegedy, 2015) to reduce internal covariate shift to accelerate deep network training. *Figure 12* shows flatten operation.

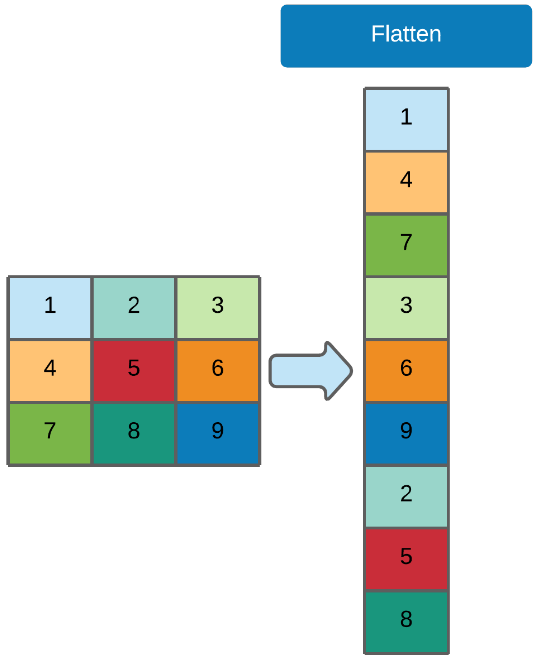

Figure *12* :Flattening layer

2.4 Machine Learning for Image analysis

As described above, Deep Learning (DL) is a sub section of Machine Learning (ML). DL is an AI technique that mimics the human brain in processing data and creating patterns. Deep Learning algorithms rely on multiple levels of interconnected neurons and have formed the computational backbone of innovative solutions across different scientific domains from traffic management to digital health (Iqbal, R, et al., 2019). As demonstrated by various studies, Deep Learning can achieve superior classification and forecasting performance in various Machine Learning tasks. Its unique characteristics make it an excellent candidate solution for computationally intensive tasks, such as image recognition, big data analytics, and real-time sensory data use (Chen & Lin, 2014) (Iqbal, R, et al., 2019). Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are some of the most popular Deep Neural Network architectures. CNN and RNN, as typical deep neural networks, are structured in consecutive layers, where each of the layers learns more complex representations of hidden features in the data under investigation. Training is conducted by iteratively adjusting the network's weights to achieve the optimal value for a cost function representing the goodness of fitness of the model to the data (Iqbal, R, et al., 2019). RNN and CNN have been used successfully on Machine Learning tasks. Demonstrative examples are the work by (Huijuan &

Jiping, 2019), where a 1D CNN was used to extract features which were later fed into an SVM classifier, and the work by Lee and Cho, where a 1D-CNN that utilized accelerometer data was proposed for human activity recognition (Huijuan & Jiping, 2019), (Lee, et al., 2017).In the previous works, the applied deep neural networks could outperform other Machine Learning techniques and effectively support crucial Machine Learning tasks such as feature space reduction (Huijuan & Jiping, 2019) , (Lee, et al., 2017).

Deep Learning's potential in object detection, classification and other Machine Learning tasks has been used successfully in recent research and towards pothole detection purposes. Pereira et al., (2018) presented a solution for pothole detection that utilized convolutional neural networks. The solution used image data from different places and under different weather conditions and demonstrated excellent performance (Pereira, et al., 2018). In (Varona, et al., 2019)'s research, Deep Learning-based approaches were used to monitor the road surface and identify potholes. The researchers investigated the application of different deep neural networks, including convolutional and Long Short-Term Memory (LSTM) networks. Their results showcased Deep Learning approaches to achieve excellent classification accuracy in the mentioned tasks (Varona, et al., 2019). (Ukhwah, et al., 2019) utilized different deep neural network architectures to support automatic assessment of the road surface and detect potholes. Their experiments resulted in reasonably high accuracy for the Deep Learning networks and low computational times to deliver the results (Ukhwah, et al., 2019).

The researchers were able to overcome the considerable challenges of real-time analysis by utilizing a CNN-based approach. Their approach achieved better performance than other conventional methods (Chellaswamy, et al., 2020)

In (Huidrom & Das, 2013), Huidrom used a charge-couple device (CCD) camera mounted on a vehicle to detect a defect on the road surface in real-time. The camera is mounted downwards to capture the road surface's images from a moving vehicle in this method. The paper suggests using road surface gloss and calculating absolute deviation with reference to the low luminance level. They assumed that the low level of luminance signal represents the road surface itself. While higher-level represent the reflection. The paper calculates the deviation of luminance predicts road condition.

In (Bouilloud, et al., 2009), Bouilloud used the Interactions between Soil, Biosphere, and Atmosphere (ISBA)-Route/''Crocus'' model to predict France's road conditions. This model depends on short term forecasts of meteorological and long-term surface conditions simulation using spatialize meteorological data. Simultaneously, the latest computer vision and computational powers of the latest computers and GPU facilitated complex computation and Deep Learning to automate the detection of defects in the road surface. Deep neural networks (DNN) have gained popularity in the field and are being used in various areas. A large data set is required to train deep neural networks-based models. Training such models on CPU could be time-consuming and may not be financially viable. However, GPU availability and parallel processing computational power have now made it affordable and accessible to develop such deep neural networks model for the business world (Zou, et al., 2012) (H.OliveiraandP.L.Correia, 2014), (S. Mathavan, 2015), (R. Medina, 2014), (Zhao, 2010), (Correia, 2013). Convolution Neural Networks has been successfully used to recognise images (Farabet, et al., 2013), (S. Zhan, 2015), (Hariharan, et al., 2014), (Krizhevsky, et al., 2012), (Liu, et al., 2015).

Steinkraus, et at discusses the importance of using GPU for training Deep Learning models (Steinkraus, et al., 2006), The researchers demonstrated how they trained one of the largest Convolution Neural Networks to date on the ImageNet dataset to achieve better performance. The paper developed an optimise implementation of 2D Convolution Neural Networks. In this paper, the authors prepared data sets of images by changing their size to 256x256 and trained the networks on the pixels' raw RGB value. The paper also discussed the importance of activation functions and training time taken by them. The paper discussed the local response to normalization and used overlap pooling before discussing how the authors trained networks on multiple GPUs to reduce training time. (F. Farnood Ahmadia, 2008) discusses neural networks' application in an automation road extraction and vectorized high-resolution images obtained from a satellite. The network used in the experiment had the same number of input neuron as an input parameter, and the output had only one neuron. The experiment used the backpropagation method with images size 500x500 pixels.

Shaoqing (2015) has discussed how the region proposal algorithm and region-based Convolution Neural Networks(R-CNN) have contributed to object detection (Shaoqing Ren, 2015). However, the region proposal algorithm could be time-consuming and not economical. The paper proposed a change in the algorithm where proposal computation is nearly free, given the detection network's cost. The paper presents a method that shares convolutional layers with state-of-the-art object detection networks. The paper presents the object detection system called faster R-CNN which has two modules. The first module is fully connected to Convolution Neural Networks that propose regions, and the second module is the fast-CNN detector. A fast

R-CNN network takes an entire image and a set of object proposal as input. The networks produce a convolution feature map with several convolution and max-pooling layers. Then a fixed-length feature vector for each object is produced from the feature map. Later each feature vectors are fed into a sequence of fully connected layers that produces two outputs. The first layer produces SoftMax probability estimates over K objects classes, and the second layer produces four real-valued number for each of the K object classes. Each set of 4 values encodes refined bound-box positions for one of the K classes.

Kaiming (2014) discusses Convolution Neural Networks requirements and highlighted how a fixed size image may reduce the accuracy of the images or sub-images of an arbitrary size/scale (Kaiming He, 2014). The paper proposes another pooling strategy, spatial pyramid pooling, to accommodate flexible size images in the networks. The networks' structure, called SPP-net, generates a fixed-length representation regardless of images size. The paper mentioned having fixed images input images in the networks comes from the fully connected layers, not from Convolution Neural Networks layers which can generate feature maps of any size using sliding window operation. The proposed SPP layers sit on top of the convolution layer and pool the features to generate fixed-length outputs fed into FCC layers or any other classifier. SPP uses multi-level spatial bins. Multi-level pooling is robust to object deformation. The proposed SPP can pool features extracted at variable scales due to the flexibility of input scales. The proposed method -SPP- is independent of the convolutional network architectures and has been used with four different architectures to show how it can improve those architectures' efficiency. The proposed multi-level pooling improves accuracy because multi-level pooling is robust to the object deformation and spatial layout variance. The research discusses how multi-size training improves accuracy and full image representations improve accuracy. The paper also discussed how combining two models trained with different hyperparameters improves object detections.

Tsung (2017) discusses the Feature pyramid networks for object detection. As it becomes challenging to recognize objects at different scales, the feature pyramids are scale-invariant. However, FPN has been avoided in recently Deep Learning methods as it is memory intensive and time-consuming. The paper discusses why featuring each level of an image pyramid has limitations and impractical in a real application. It discusses the infeasibility of using an image pyramid to training end to end deep networks. It requires intensive memory and time and creates inconsistent training and tests time references. The paper focuses on the sliding window prospers(region proposal Network, RPN) (S. Ren, 2015) and region-based detectors (Girshick, 2015). The proposed architecture leverages the pyramidal shape of convolution Networks features while creating a feature pyramid by combining low resolution, semantically robust features with high resolution, semantically weak features via top-down pathway lateral connections. This method produces a feature pyramid with rich semantics at all levels and is built quickly from a single image scale. In this paper, handcrafted image features have been replaced with automatically computed features by convolutional neural networks. Convolution neural network is robust to the variant in scale and can facilitate recognition from features computed on a single input scale. The method shows improvement over several strong baseline networks (Tsung-Yi Lin, 2017).

Gao et al. (2012) discusses how the Fisher linear discriminant analysis (FLDA) methodology can significantly increase image recognition. (Gao, et al., 2012) However, FLDA, on the other hand, ignores the variance between data points from the same class. The variance of nearby data from the same class, which characterizes the most significant trend variation modes and contributes, causes the Fisher discriminant criteria to be unstable. The paper proposed an improved Fisher discriminant criterion (EFDC). EFDC specifically acknowledges intra-class heterogeneity and integrates it into the Fisher discriminant criteria to construct a stable, effective dimensionality reduction function. EFDC obtains a subspace that better senses discriminant structure while still preserving the heterogeneity of patterns, resulting in consistent intraclass representation. The experiment used the six discriminant approaches on four image databases and concluded that this proposed method indicates a significant increase over the results of FLDA-based approaches.

Hui et al. (2018) discusses Convolution Neural Networks' weak performance under limited sample sizes and the proposed face recognition approach based on Convolution Neural Networks and the Fisher criterion (Hui & Yu-jie, 2018). In this method, a discriminant metric parameter is applied to the error's cost function to improve network classification. The facial features are then extracted using the updated convolution neural network. Finally, support vector machine's benefit in limited sample size, nonlinearity, and large dimension is used to characterize the extracted functions. The experimental findings indicate that the face recognition algorithm built on the Fisher neural network and SVM can obtain decent results with fewer samples. The method shows that when the number of samples is limited, the proposed approach outperforms other recognition quality methods. When the number of samples per class exceeds 8, the identification rates of various approaches are almost identical.

Shi et al. .(2016) discusses the issue of road cracks and why automatic road crack detection is essential. The paper proposes a novel crack detection approach that takes advantage of the hierarchical knowledge contained in cracks (Shi, et al., 2016). This system is divided into three parts: In the first part, they broaden the feature set of typical cracks detection by integrating the integral channel functionality into the method. These extracted features from multiple levels and orientations allow us to re-define representative crack tokens with more organized detail. In the second part, random structured forests are added to manipulate such structured knowledge, allowing a preliminary crack detection result to be obtained. In the third section, they suggest a new crack descriptor based on the statistical properties of tokens. This descriptor will characterize cracks of any topology. A classification algorithm (KNN, SVM, or One-Class SVM) is used to differentiate cracks from sounds efficiently. The paper used across data sets shows that the proposed method's accuracy and stability are comparable or better.

Chen et al. . (2018) talked about the importance of information about road terrain for intelligent vehicles. This paper proposes a method for recognizing road terrain using photographs of road surfaces. They collected images of the road surface by mounting a camera on the top of a frame of the vehicle's back tray. The camera was facing downward, and the camera's vision was adjusted to capture just the road surface. The images were collected under variable circumstances of different natural driving condition and environments. The following stage features were extracted from the images and fed in the support vector machine (SVM) classifier. This study's accuracy was above 90% except for the asphalt road, which has proven difficult to classify and had an accuracy of only 65.6%. The results show that this method is suitable for terrains with a distinct texture. However, the technique could not identify road when there was variation in light or when the vehicle's speed was different (Chen, et al., 2018).

Roychowdhury et al., (2018) discussed how Active protection technologies and self-driving vehicles could profit significantly from a real-time prediction of drivable surface conditions so driving styles can be tailored to the following road conditions. The paper discusses the direct and indirect method of road friction estimation (RFE). Before real-time road friction estimation, sensor data and road tire friction were correlated to find road friction. It then discusses the literature that has used neural networks, Support Vector Machine, Naïve Bayesian and Random Forest, to predict road friction. This study proposed a two-stage method to classify road surface condition using imagery data and then segmentation the road for REF classification. The first stage used Convolution Neural Networks to extract feature from the images to train the model and in the next step, it classified the surface into three categories.

The method achieves 94-99% accuracy in the first stage and 89% in the RFE classification stage. These accuracies are better than the other Neural Networks based method, which is mentioned in this study (Roychowdhury, et al., 2018).

Qiu, et al. .(2020) discusses why a good road surface is essential and the difficulties in effectively detecting abnormal road surface. It proposes a K-mean algorithm and Gaussian background model to identify abnormal road surface and give their geographical location. The data was collected using a smartphone and customised mobile application. The results have been discussed in absolute number and there is no mention of percentage accuracies of any comparison with other results (Qiu, et al., 2020). (Banica, et al., 2017) Discusses hardware and software to collect data from multiple sources, including laser-based imaging, proximity sensor and environmental images and they correlate this information using odometry and geolocation.

Jang, et al. (2015) discusses the existing methods of road surface monitoring and how these methods depend on reporting from drivers. This method has a significant lag, so it is impossible to access road surface data in real-time. This paper suggested collecting accelerometer and GPS data using a microcomputer. They used a supervised Machine Learning technique at the backend – multiplayer feed-forward neural networks to classify the collected data into three classes. The proposed classification method's accuracy is 83.2%, 79.7% ad 91.9% for the respective class (Jang, et al., 2015).

Xia (2018) used an HD camera to collect images of road surface and classify those images into five damaged classes (Xia, 2018). They used weakly supervised labelling to make big data set and later used supervised Machine Learning -Convolution Neural Networks method for classification. Mandal, et al. (2018) proposed a Convolution Neural Networks-based method that uses YOLO v2 Deep Learning framework. They used images collected using a smartphone (Mandal, et al., 2018). Wiratmoko and Syauqi (2019) proposed a device to collect road potholes and used Convolution Neural Networks for pothole detection. They collected videos which were later converted into images and then fed into a convolution neural network. The study has 92.8% accuracy (Wiratmoko & Syauqi, 2019). In (Islam & Sadi, 2018), Islam & Sadi proposed a Convolution Neural Networks-based method to classify road potholes. They used publicly available databases, which has a total of 379 images. The experiment was run for 30 epochs and obtained 97.12 % and 97.3% testing and training accuracy. Medvedev & Pavlov (2020) proposed Convolution Neural Networks based road surface marking recognition method. This paper focuses on developing a cost-effective computational method that can be operated even on a smartphone. The study used a pre-trained neural networks model and customised for this study. The study used Faster RCCN for classification. The study achieved an accuracy of 98%. (Medvedev & Pavlov, 2020). Chen & Wang (2019) discussed that LeNet-5 Convolution Neural Networks is not accurate when classifying complex images (Chen & Wang, 2019). The study proposed adding new layers to deep the networks to extract low-level features more effectively. Later the low-level features and high-level features were combined. The Adam optimizer was used to adjust and update the network's parameters. They used CiFar-10 and Fer2013 publicly available dataset to run the experiment. The study achieved 98.48%, which comparable better than LeNet-5's.

Ukhwah, et al. (2019) utilized different deep neural network architectures and images to support automatic assessment of the road surface and detect potholes. Their experiments resulted in reasonably high accuracy for the Deep Learning networks and low computational times to deliver the results (Ukhwah, et al., 2019).

As discussed above, computer vision and Deep Learning are being used to classify objects. Convolution Neural Networks have been applied for image classification (Krizhevsky, et al., 2012), (Luo, et al., 2016) and object detection (Krizhevsky, et al., 2012), (Wang, et al., 2015). The image-based method is cost-effective in comparison to the 3D laser scan method. However, an image-based method is sensitive to environmental factors such as light, shadow, water etc.

2.5 Machine Learning for Sensory Data Analysis

In recent time, smartphone and Internet of Things (IoT) has emerged rapidly and have become an integral part of day-to-day life. According to a report by Ericsson (Ericsson, 2016) there will be 31.4 billion smarts connected IoT devices by 2023. These intelligent devices are generating lot of valuable data which can be used to solve many important issues such as road pothole detection. The focus of this section is to review previous research work which have used Machine Learning and sensory (accelerometer) data for pothole detection.

The method to detect road pothole using sensor data comes under vibration-based method. This method is cost-effective, requires small storage and can be used in real-time (Kim & Ryu, 2014). The vibration-based methods can broadly be divided into three categories:

(1) Threshold-based methods,

(2) Dynamic Time Warping (DTW),

(3) Machine Learning methods.

Artis et al (2011) discussed the vibration method to detect potholes in their work. Data samples were collected using a customised application and later, detection algorithms Z- thresh, and Z-peak were applied to find potholes (Artis, et al., 2011).

In (Chen & Lu, 2013), Chen &Lu designed and developed a device to collect accelerometer data. The device collected accelerometer data on three axis-X, Y and Z, along with GPS information. The study discussed why the Z-Peak method and why a uniform threshold cannot classify road surface. The study prosed the i-Gaussian Mixture Model (GMM), which accommodated the variability of speed.

Gunawan, et al. (2015) have proposed an accelerometer data and filter base method to classify road surface. The data set for the study was prepared by driving a car to collect accelerometer data. During the data pre-processing, the data sample that was not good for the experiments was dropped, and new data were collected. Later, a threshold was applied on the Z's axis data to classify road surface (Gunawan, et al., 2015).

Yi, et al., (2015) proposed crowdsourcing using a smartphone to collect sensor data. The next stage used a clustering algorithm – DENCLUE- to extract road anomalies and index those. (Yi, et al., 2015). Chen, et al (2016) designed a specially designed device to collect images and sensor data of the road surface. They have used Iterative Closest Point (ICP) to identify potholes. The accuracy of the model is in the range of 90% (Chen & et al., 2016).

Harikrishnan, et al. (2017) proposed a method in which accelerometer data is collected using a smartphone and later Gaussian model-based and x-z ratio filter is applied to detect road anomalies. The data was collected when the vehicle was running at a speed of 15-20 km/h. The estimated error was in the range of 5.08% to 61.93% (Harikrishnan & Gopi, 2017).

Singh et al (2017) used accelerometer sensor data to detect anomalies using the DTW method. The method produced accuracy in the range of 88% and was not sensitive to speed (Singh, et al., 2017). Mohan et al. (2008) proposed a two detectors method to detect bumps and potholes. The proposed method is sensitive to the speed and was conducted at 25 km/h (Mohan, et al., 2008).

Alqudah & Sababha(2017) proposes the use of the embedded sensor in a smartphone to collect gyro rotation data. They used variability in the gyro rotation data to detect abnormality in the road surface. They used Google map to find the distance of abnormalities to validate their results in the first stance. Later they used Dynamic Time Warping (DTM) to find road abnormalities (Alqudah & Sababha, 2017).

The threshold-based methods detect anomalies when there is a change in amplitude or some of the signal's other properties compared to a specified threshold value. Dynamic Time Wrapping (DTW) measures similarities between two sequences which may vary in space and time (Sattar, et al., 2018). In the recent work by (Chellaswamy, et al., 2020), a system that utilized ultrasonic sensory data was proposed for identifying humps. In (Davide & Alessandro , 2012), Davide & Alessandro described how to make a standard human activity recognition dataset and published a dataset named ‘Activity recognition using smartphones dataset’. The same dataset was used in (Davide & Alessandro , 2012) to recognise human activities using SVM.

In (Ikeda & Inoue, 2018), Ikeda &Inoue discussed their motivation to help people during natural calamity. They proposed collecting accelerometer data from pedestrians’ smartphone and sued the supervised Machine Learning method -Support Vector Machine to decide whether the surface is flat. Before using the SVM classifier, another used statistical method to extract feature from the sensor data. After classification of the surface, nodes were generated, and a map of the safe route was predicted.

Dey,et al. (2019) proposed a sensor data-based method and Machine Learning to classify road surface. They had developed a smartphone application to capture sensory data. The study has used Best Fist, Ranker and Greedy Stepwise to optimise feature selection. Later the study has used Support Vector Machine, Random Tree and Random Forest to do the surface classification. The results have been discussed with respect to the features optimisation method. The best result has been noticed for the Random Forest approach (Dey, et al., 2019).

Varona et al., (2019) used Deep Learning method and sensor data were used to monitor the road surface and identify potholes. The researchers investigated the application of different deep neural networks, including convolutional and Long Short-Term Memory (LSTM) networks. Their results showcased Deep Learning approaches to achieve excellent classification accuracy in the mentioned tasks (Varona, et al., 2019).

EL-kady, et al, (2019) proposed a method to detect road anomalies on the road of Egypt. They developed an application for android smartphone and collected accelerometer data. In the next step, they used the K-Mean clustering technique to decide whether the road has any anomaly. Later they used Support Vector Machine to classify an unseen road surface (EL-kady, et al., 2019).

Chao et al., (2020) discussed the use of a Machine Learning approach for road pothole detection using smartphones. The paper discussed data processing and Machine Learning classification methods, such as logistics regression, Support Vector Machine (SVM), and random forest to detect potholes. In the proposed method, features from collected data were extracted, and a Machine Learning classification was used to detect potholes. This paper did not use neural networks, arguing that it requires a large set of data and does not guarantee higher efficiency (Chao , et al., 2020).

In the recent work by (Chellaswamy, et al., 2020), a system that utilized ultrasonic sensory data was proposed for identifying bumps. The researchers were able to overcome the considerable challenges of real-time analysis by utilizing a Convolution Neural Networks -based approach. Their approach achieved better performance than other conventional methods and more straight forward Machine Learning techniques (Chellaswamy, et al., 2020).

The methods review above have used sensory data with either statistical methods or Machine Learning method to detect road pothole. There are many researchers who have used fusion method to combined sensor data to sensor data or have combined used images along with sensory to detect road pothole. The following few reviews cover those literatures.

Z (2010) proposed a fusion method to combining three-axis accelerometer data and Fast Fourier transform (Z, 2010). Following that a Principal Component Analysis was applied to reduce the dimension of the fusion features. Finally, the Support vector Machine classifier was used to classify the dataset. This research achieved an accuracy of 89.89%. The data was collected on a mobile phone. The number of participants in the experiments was sixty-seven, and they had performed 17 different gestures.

Chen,et al. (2016) proposed a vision based multi-sensor system to monitor road surface. This paper used an RGB camera to capture images and a custom-designed sensor to collect accelerometer and GPS data. After collecting the data, they applied the method to collect the relative depth of the normal road surface and the area with a defect. The depth of defective area is calculated using the relative value of the histogram. The accuracy of the model is in the range of 90%. The product's cost was over $ 7000, which is a lot compared to a normal smartphone (Chen, et al., 2016).

In (Gueta & Sato, 2017), Gueta &Sato collected accelerometer data, GPS and time details by mounting an accelerometer data recording device under the driver's seat. They also recorded videos of the road to observe anomalies and label accelerometer data accordingly. The study used, Support Vector Machine to classify the road surface. The study achieves 77% classification accuracy.

Nobis, et al. (2019) proposed a fusion architecture that uses data from Radar and Camera for object detection. They discussed the difficulty to combine data from radar and camera as radar data has no information about height of the object. They made the corresponding value zero to address this issue and later incorporated the radar data with camera data. They used RetinaNet with a VGG backbone. They used a publicly available dataset called Nuscense dataset. The experiment achieved accuracies of 43.95% and 57.50% (Nobis, et al., 2019) .

In (Zhou, 1994), Zhou proposed a method to fusion images from different sources and of various sizes. This method reduces noise in the output images. The hierarchical fusion method which is based on human vision system was used in this study.

However, the vibration method fails to differentiate among potholes and other forms of anomalies, such as hinges and joints on the surface (Kim & Ryu, 2014).

Table 1:The list of works and their area of work

Reference	Used for	Input	Data collection	Method	Classifier	Result
(Gao, et al., 2012)	Image recognition	Images	Public	Principle Component Analysis	Nearest Classifier	Improved accuracy
(Hui & Yu-jie, 2018)	Facial Detection	Images	Custom	Machine Learning	Support Vector Machine	Improved accuracies
(Shi, et al., 2016)	Road crack	Images	Custom	Machine Learning	Support Vector Machine	Comparable and Better
(Chen, et al., 2018)	Road Terrain	Images	Custom	Machine Learning	Support Vector Machine	Improved except in asphalt surface
(Roychowdhury, et al., 2018)	Road friction	Images	Custom	Machine Learning	Convolution Neural Networks	Improved accuracy
(Qiu, et al., 2020)	Road surface	Images	Custom	Machine Learning	K-Mean and Gaussian	---
(Alqudah & Sababha, 2017)	Road Surface	Sensor Data	Custom and	Statistical method	Dynamic Time Warping	---
(Jang, et al., 2015)	Road Surface	Sensor data	Custom	Machine Learning	Feedforward Neural Networks	Good accuracy
(Xia, 2018)	Road Surface	Images	Custom	Machine Learning	Convolution Neural Networks	--

(Mandal, et al., 2018)	Road crack	Images	Custom	Machine Learning	Convolution Neural Networks, YOLOV2	Good accuracy
(Wiratmoko & Syauqi, 2019)	Road Potholes	Images	Custom	Machine Learning	Convolution Neural Networks	Good accuracy(92.8%)
(Islam & Sadi, 2018)	Road Potholes	Images	Public databases	Machine Learning	Convolution Neural Networks	97.2% and 97.3%
(Chen & et al., 2016)	Road Surface	Sensor data,	Custom	Statistical method	Displacement method	90%
(Yi, et al., 2015)	Road Surface	Sensor data	Custom	Machine Learning	Clustering algorithm and Indexing	-
(Harikrishnan & Gopi, 2017)	Road Surface	Sensor data	Custom	Statistical method	Filter	Error in the range of 5.08% to 61.93%
(Ikeda & Inoue, 2018)	Road Surface	Sensor data	Custom	Machine Learning	Support Vector Machine	In the range of 30% and 86.61%
(Dey, et al., 2019)	Road Surface	Sensor Data	Custom	Machine Learning	Support Vector Machine	92%
(Medvedev & Pavlov, 2020)	Road Surface	Images	-	Machine Learning	Convolution Neural Networks	99%
(Chen & Wang, 2019)	Image classification	Images	Public	Machine Learning	Convolution Neural Networks	98%

(Gunawan, et al., 2015)	Road Surface	Sensor Data	Custom	Statistical Method	Filter	-
(Gueta & Sato, 2017)	Road Surface	Sensor data	Custom	Machine Learning	Support Vector Machine	77%
(Chen & Lu, 2013)	Road Surface	Sensor data	Custom	Statistical Method	Filter	92%
(EL-kady, et al., 2019)	Road Surface	Sensor data	Custom	Machine Learning	Support Vector Machine	>95%
(He, 2010)	Hand Gesture	Sensor Data	Custom	Machine Learning	Support Vector Machine	>85%

2.6 Chapter Summary

This chapter has reviewed state of the art Machine Learning approaches. Particularly this chapter has discussed Deep Learning approaches and its application for imagery and sensory data processing. The perceived limitations of these approaches are highlighted in this chapter.

This chapter has also briefly covered the concept of Machine Learning and different types of Machine Learning algorithms such as supervised and unsupervised. After a brief discussion of Machine Learning, the chapter discussed Deep Learning and described the architecture of Convolution Neural Networks in detail covering various layers and works of these layers. Later in this chapter, a comprehensive literature review is presented to understand the use of computer vision, sensors and Deep Leaning in image classification and in road surface monitoring. The literature review has given the clarity of work done in the area of road surface monitoring to detect pothole. The perceived limitations of the existing work have been unfolded which will further be investigated and addressed in the proposed research.

As discussed above, the Deep Learning method along with Computer Vison method is being widely used for road surface monitoring. However, the computer vision method to detect road anomalies fails to differentiate among pothole, muddy patch and shallow puddle-which may not be deep enough to be classified a pothole. The vibration method fails to differentiate between a pothole and other anomalies (Kim & Ryu, 2014). Also, to the best of the author's knowledge there is no research which has used Deep Learning to identify road pothole using sensor data. To overcome the weakness of computer vision and vibration system, this Thesis will consider developing a Deep Learning fusion model to process two inputs: image and accelerometer data to detect road potholes.

The next chapter will discuss the proposed cloud-based architecture and the research methodology in detail.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

CHAPTER 2 covered a review of the existing approaches in image processing, object detection and classification using various Machine Learning approaches and various data types such as imagery and sensory data. Based on the review; it is found that the Deep Learning method can deliver good results in the area of pothole detection. However, after extensive research, it was noticed that there are some flaws in computer vision and sensory data application methods. The computer vision method, which uses imagery data to detect road anomalies, fails to differentiate between pothole and puddle. The vibration method, which uses sensory data, fails to distinguish between a pothole and other anomalies such as hinges. In order to overcome the limitations of computer vision and vibrations systems as identified in the previous chapter, this thesis proposes a fusion model that utilizes two inputs: image and accelerometer data to detect road potholes.

The rest of the chapter is organised as follows. Section 3.2 will discuss the proposed cloud-based model to report road potholes and share information on the updated road pothole database with the stakeholders. Section 3.3 will describe the methodology of the Deep Leaning model to detect road pothole using imagery data. The section will mention various steps involved in imagery data collection and the design of the Deep Learning model. Section 3.4 will describe the methodology to collect sensory data and design a Deep Learning model that uses sensory data to detect road potholes. Section 3.5 will list the steps involved in designing and training the fusion model.

3.2 Proposed Cloud based Architecture

The proposed cloud-based architecture to report pothole and information sharing is shown in Figure 13. It shows how a pothole is reported to the authority automatically. When a road user's vehicle approaches a road pothole, the user's smartphone that is mounted on the vehicle's windshield will capture images and sensory data. These data will be uploaded to the cloud server. The data will be processed at the back end, and the road pothole database will be updated in real-time. Any road users can access the database of road pothole. Also, the subscribed user will get an alert notification when they are close to the pothole.

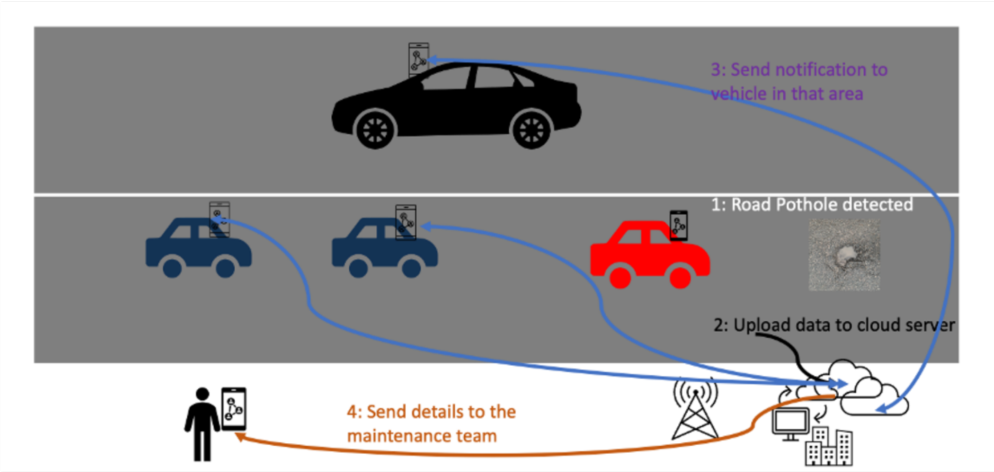

Figure 13:Cloud based road pothole information sharing architecture

Figure 14 shows all the steps involved in the end-to-end process of spotting a new pothole including reporting it; the process involved at the backend to detect a pothole; updating the cloud-based database; and finally sending a notification to the interested stakeholders.

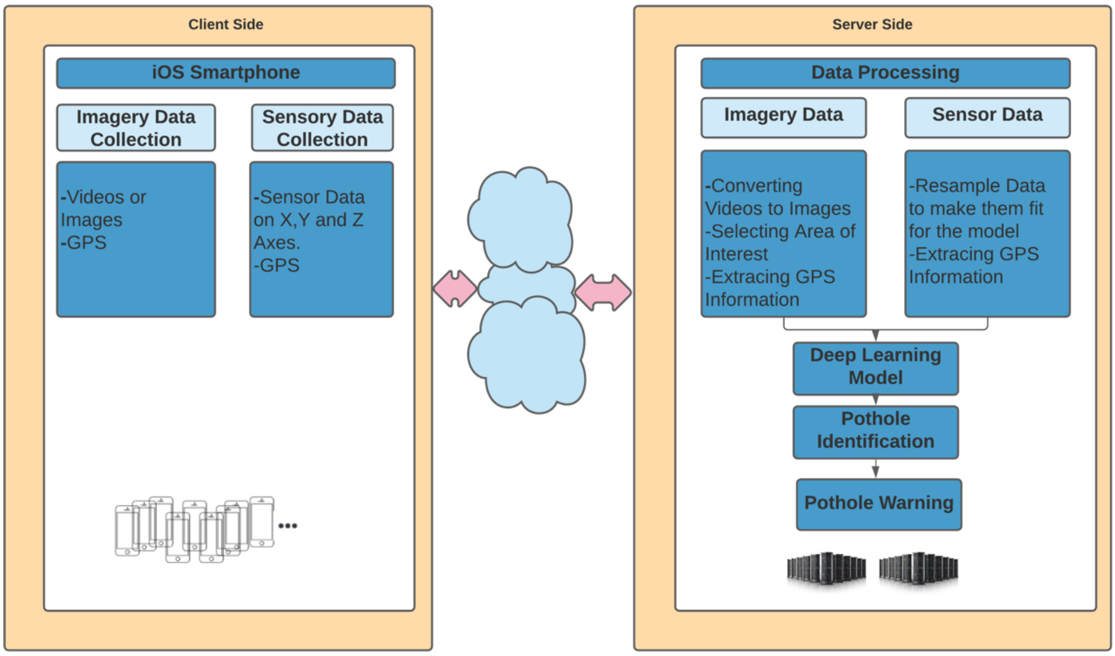

Figure 14:The architecture of the proposed cloud-based road pothole detection method

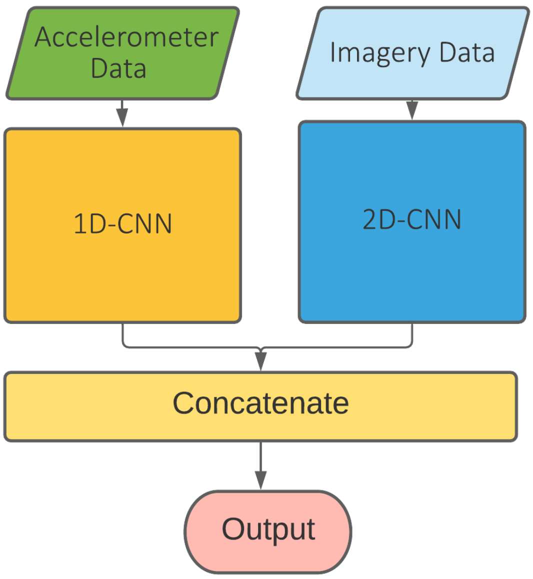

Figure 15:Proposed deep learning fusion model

As discussed earlier in this section, this research is aimed at developing a fusion model of one-dimensional Convolution Neural Networks (1D-CNN) and two-dimensional Convolution Neural Networks (2D-CNN), as shown in Figure 15. The final fusion model was developed after experimenting with the 2D-CNN model and the1D-CNN model separately. Figure 16 shows how the research project has progressed after conception to obtain the final fusion Deep Learning Model. The first experiment was aimed at designing a 2D-CNN model to identify road potholes using imagery data. The methodology for this experiment is discussed in section 3.3. and the results of experiment -1 are presented and discussed in section 5.2. After experimenting with the 2D-CNN model and imagery data, then a 1D-CNN model was designed and trained with sensory data. The methodology for experiment-2 is described in section 3.4.

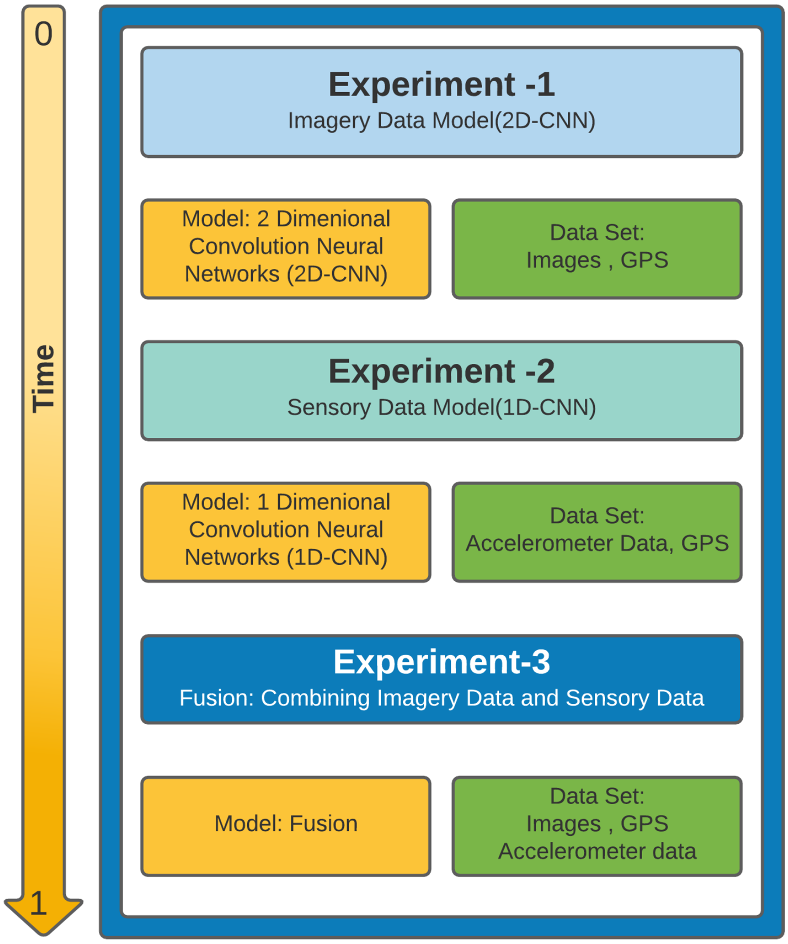

Figure 16:Research project's stages

The details of experiment-2 and the corresponding results are mentioned in section 5.3. After analysing the results from Experiment-1 and Experiment-2, the final Experiment-3 was conducted to design and test a fusion model. The methodology for the fusion model is explained in section 3.5 . The details of Experiment-3 and results are discussed in section 5.4.

After a brief investigation, it was observed that there are a few datasets available, but none, to the best of the author's knowledge, are derived from the UK. Also, those data sets were prepared for the specific projects and are not suitable for the undertaken research. The imagery data and sensor data were both collected using an iOS smartphone. The data set was augmented to enhance the quality and size. The preparation of the data set is discussed in CHAPTER 4. Once the data was ready to be used in the Convolution Neural Networks models, a few 2D-CNN model and few 1D-CNN models were used to run the experiments. In the 2D-CNN models, different number of hidden layers, different hyperparameters and different images sizes, but same for a dataset, were used to obtain an optimal model. In the 1D-CNN models, different number of hidden layers and different hyperparameters were used to run the simulations to obtain an optimal 1D-CNN model. The 2D-CNN model and 1D-CNN model with highest accuracy, highest Average Precision Rate and highest Average Recall Rate were used to make a fusion Deep Learning model. The fusion Deep Learning Model took two inputs, imagery data and sensor data, and trained. The following sections discuss the methodology in detail.

3.3 Deep Learning Method for Imagery Data

As shown in Figure 17, the first stage was to collect imagery data. A smartphone was mounted on the windscreen of the vehicle to record videos and take images. The videos were later changed to photos using a Python script. The images were investigated carefully to remove unwarranted images and then they were divided into two classes: pothole and no pothole. Once the imagery data set was ready, then in the next stage, a Deep Learning model was designed. The simulations were run using various combinations of hidden layers and hyperparameters to obtain a satisfactory model.

Figure 18 shows the steps involved in identifying a road pothole using imagery data. The imagery data was collected, and the area of interest was selected before feeding the imagery data into the trained 2D-CNN model. The output of the model was imagery data with identified road pothole and their geographical location coordinates.

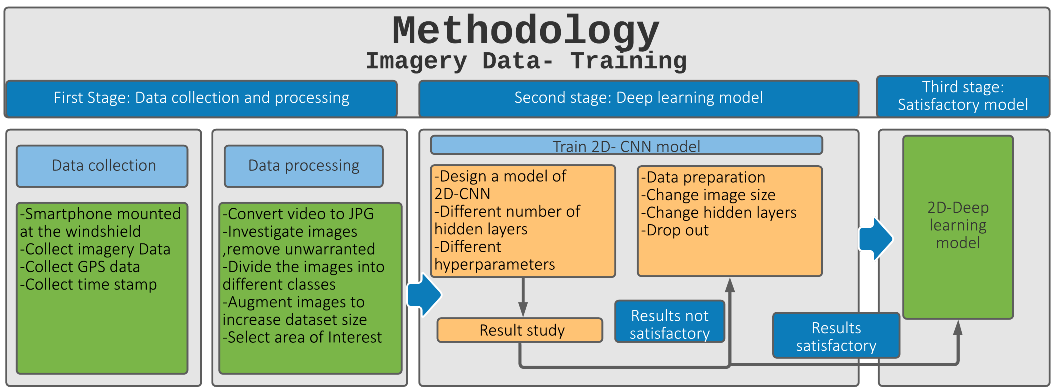

Figure 17 : Methodology to train 2d-cnn model on imagery data

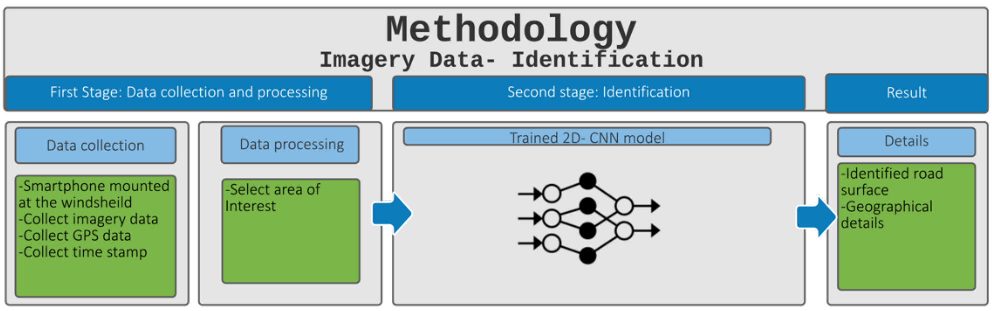

Figure 18: Methodology to identify pothole using imagery data with 2d-cnn model

3.4 Deep Learning Method for Sensor Data

Figure 19 describes the steps used to train a one Dimension Convolution Neural Networks (1D-CNN) on sensor data collected using a smartphone. Once the sensory data set was prepared, as described in section 4.3.1, it was fed in a 1D-CNN model. The various combinations of hidden layers, kernel size, and drop out were tried to obtain an optimal 1D-CNN model. Figure 20 shows the steps to identify a road pothole using sensor data. The sensor data was collected and resampled, and then fed in the trained 1D-CNN model. The output of the 1D-CNN model was an identified road surface with their geographical location details.

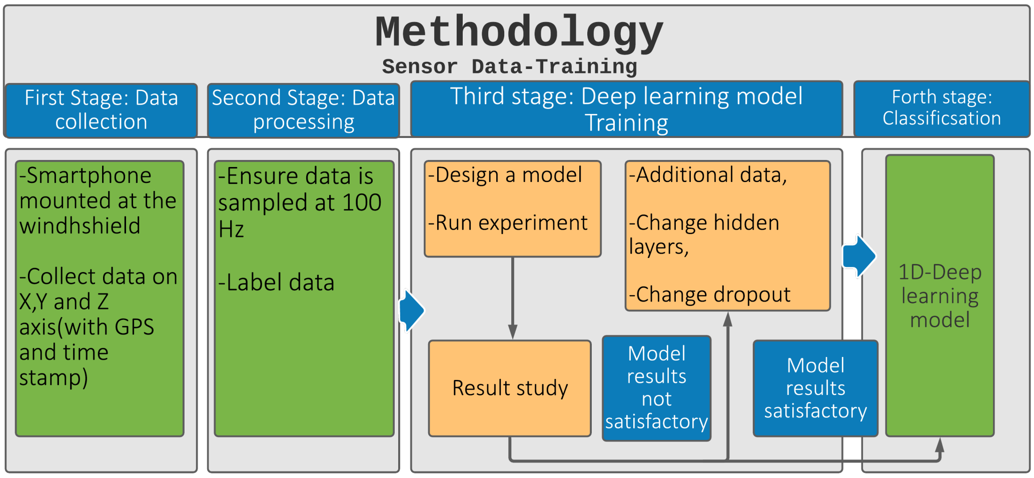

Figure 19:Methodology to train 1d-cnn model on sensor data

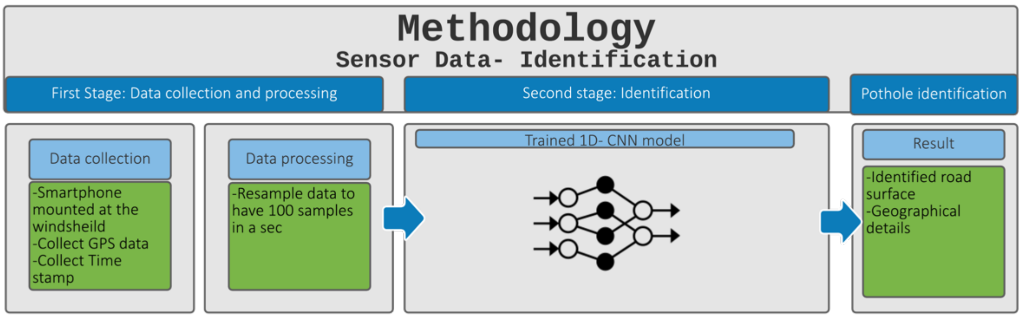

Figure 20:Methodology to identify pothole using sensor data with 1d-cnn model

3.5 Fusion Method

The optimal 2D-CNN and 1D-CNN models identified in the previous experiments were used to start designing the fusion model (see Figure 21). The fusion model took imagery data and sensor data as the inputs and was simulated with various hidden layers and other hyperparameters to obtain an optimal fusion model. The data set preparation is discussed in detail in section CHAPTER 4. Once the data set was ready, it was fed into the fusion model, as shown in Figure 21 to simulate the fusion model. Figure 22 shows the steps involved to identify a road pothole using the trained fusion model.

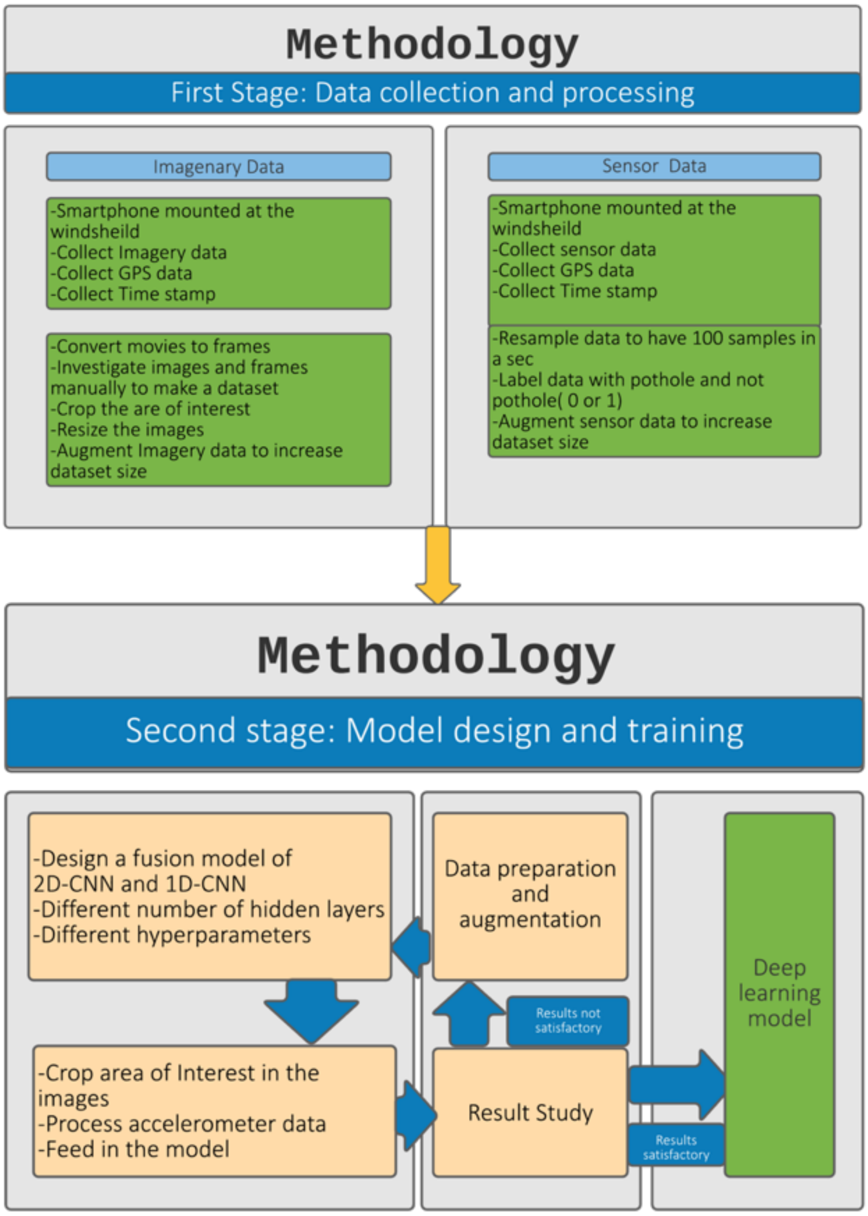

Figure 21 : Methodology to train fusion CNN model using imagery and sensor data

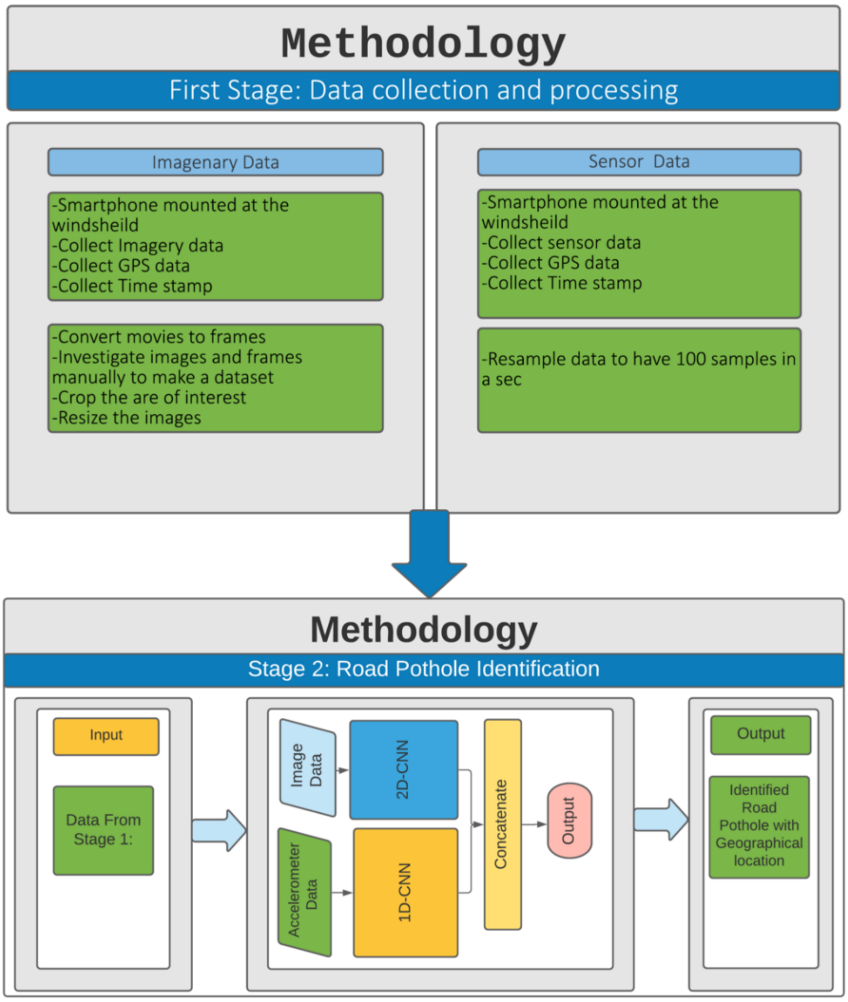

Figure 22: Methodology to identify road pothole using the fusion CNN model with imagery and sensor data

3.6 Chapter Summary

This chapter has discussed the research methodology in order to systematically utilize imagery data for two-dimensional Convolution Neural Networks (2D-CNN). The chapter has also discussed how to use the trained two-dimensional Convolution Neural Networks model for pothole detection. Furthermore, this chapter described the steps to train a one-dimensional Convolution Neural Networks model (1D-CNN) This chapter has outlined how to use the 1D-CNN model to detect pothole. Most importantly, the chapter described how to collect data and train a fusion model with sensory and imagery data. Also, this chapter has highlighted how to use the trained fusion model for pothole detection.

The next chapter will discuss methodologies for imagery and sensory data collection as well as pre-processing of data.

CHAPTER 4

DATA COLLECTION AND PRE-PROCESSING

4.1 Introduction

Data collection and pre-processing is a significant factor in Machine Learning and a hot subject of discussion in many communities (Roh, et al., 2021). Data collection has lately been a serious topic for a couple of reasons. Firstly, as Machine Learning becomes more mainstream, there are more applications that do not often have enough labelled data. Secondly, Deep Learning techniques produce features automatically, saving feature engineering costs while potentially requiring more labelled data. Data pre-processing primarily consists of data acquisition, data labelling and data preparation.

The rest of the chapter is organised as follows. Section 4.2 will describe the setup for data collection and the dependency on placement and orientation of smartphone, section 4.3 will discuss data collection, data pre-processing, augmentation methods and labelling for sensory data, 4.3 will discuss data collection, pre-processing, augmentation and labelling of imagery data. Finally, section 4.4 will present a summary of the chapter.

4.2 Data Collection

A standard data acquisition flowchart is shown in Figure 23. It checks if there is enough data to train a Deep Learning mode. If there is not enough then data needs to be collected and pre-processed before feeding into Deep Learning model.

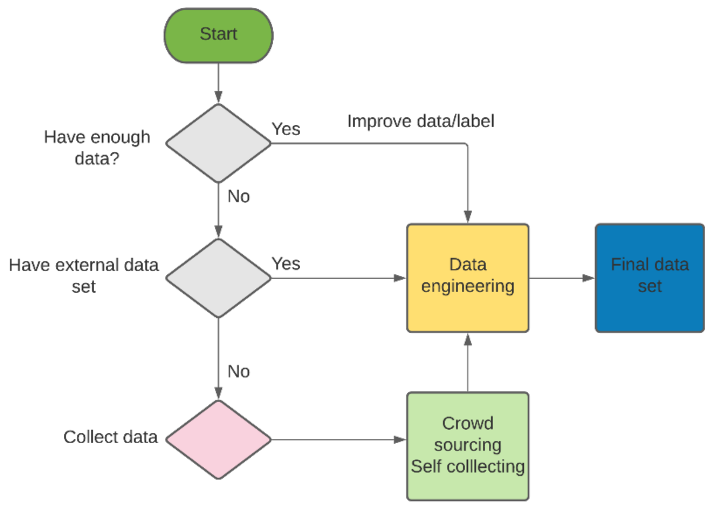

Figure 23: Data collection and pre-processing flowchart

The Imagery data has been collected using a smartphone camera (Iphone, 2019) and the sensory data has been collected using an app called Sensor Data from Wavefront Labs (WavefrontLabs, 2019) installed on the same smartphone. The collected data are stored locally at the iPhone and later transferred to the computer described in Table 9 to perform the experiments. The smartphone was securely placed at the windshield of a vehicle, as shown in Figure 24. The camera of the smartphone was used to collect videos. Simultaneously, the application installed on the same smartphone was recording 3-Dimensionsal (X, Y and Z) accelerometer data, along with the corresponding timestamp and GPS information. The sampling rate of the accelerometer was set at 100 Hz. However, due to hardware limitations, the GPS sampling rate was set to 1 Hz. The data was stored on the iOS smartphone and later downloaded on the computer Table 9 for data processing, modelling and analysis. This study's data was collected while driving on the motorway, A-roads, B-roads and in town.

It was observed that motorways and A-roads were in good condition, and most potholes were noticed on B-roads and inside the town. The pothole-data used in this study were collected while driving inside the town or on B roads with an effort to keep the speed steady at 30 miles per hour. There were variations in speed, ex: stopped at red lights or lower speed in congestion, however data with speed at 30 miles per hour were used in this experiment to keep the consistency. The time to cover the distance mentioned in Table 2 was nearly 2 hours and 50 minutes. The GPS data and image sampling rates were 1 Hz due to hardware restrictions of the smartphone which was used in these experiments. So, for each second the GPS location was unique. The first frame of the video was used, if the frame was not blur and usual, as an image. The higher sampling rate will yield better location accuracy. The sensor data sampling rate was

Table 2:Distance covered to collect data and road type

Road Type	Road Quality	Distance(miles)
Motorway	Very Good	80
A Road	Good	40
B road and in town	Not good	30

set at 100 Hz, which was later merged to make datapoint per second. Each data point represents one second which will have total 300 sensory data samples, 100 data samples on each of the X, Y and Z axis, as frequency is set at 100 Hz due to the hardware restriction, and the corresponding image of road surface. At the mentioned sampling rate, there were 7200 data points which were useable. Later, sensory data and imagery data were augmented, which is discussed later in this section, to increase the size of the data set and achieve better accuracy (Jing & Tian, 2020). Table 3 and Table 4 show the classes and the number of data samples in the respective classes in the final data set after augmentation. In the

Table 3:The final dataset for 2d-cnn model

Damage	Road Quality	Number of	Image size
DM00	Normal Road Surface	2810	256x256
DM01	Small pothole as seen from the dashboard camera	4074	256x256
DM02	Seriously damaged surface. Images captured using a handheld device(iPhone)	3694	256x256

Table 4:The final dataset for sensor model and fusion model

Damage Category	Road Quality	Number of Images	Images Size (pixels)	Number of samples of sensor data
DM00	Normal Road Surface	22344	128x128	22344
DM01	Pothole as seen from the dashboard camera	11016	128x128	11016

4.2.1 Smartphone Placement

The positioning of the smartphone on the windshield is an essential factor for achieving high accuracy in pothole detection. As discussed, an iOS smartphone was used to capture images and videos for this study. The smartphone was securely placed at the middle of the windshield's width and at the highest point as shown in Figure 24, in order to have a clear view of the road. Figure 25 shows how the smartphone camera sees the highway, and Figure 25 shows an example of a captured pothole. The same smartphone, which was securely placed at the windshield, was used to record the accelerometer data. There was no ambiguity or confusion about the smartphone's axes and the vehicle's axes. In most research efforts that were reviewed, the smartphones were mounted on the windshield or the dashboard. A limited number of researchers have studied the performance when the smartphone was kept in the glove box or the driver’s pocket. As demonstrated by the experimental results in these studies, the placement of the smartphone in the driver’s pocket or the glovebox resulted in lower detection rates.

Figure 24: Smartphone placement at the windshield

Figure 25: The road as seen from the smartphone

Figure 26: sample of a pothole

4.2.2 The Orientation of The Smartphone

Image (Figure 26) shows that a pothole is two-dimensional shape. Therefore, pothole detection will not be impacted by the placement of the camera on the windshield. Figure 27 shows the smartphone camera placement in detail. Figure 29 and Figure 28 show the smartphone and vehicle axes. However, pothole detection using accelerometer data is sensitive to the orientation of the sensors. (Yagi, 2010) and (Mednis, et al., 2011) assumed a fixed position for analysing data from the smartphone. (Chao , et al., 2020) applied Euler angles to align the orientations for their study. The smartphone in this study was placed securely upright as shown in Figure 24, at the windscreen to make sure that accelerometer data for all three axes are in synch with the axes of the vehicle

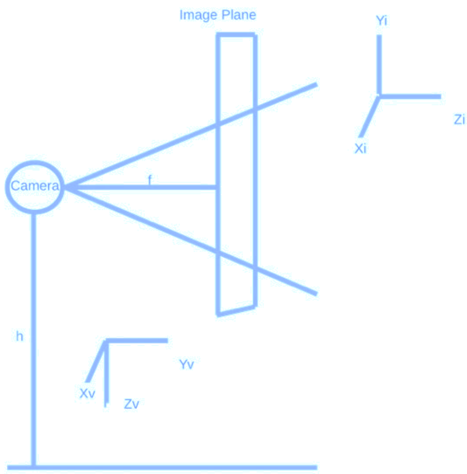

Figure 27:Smartphone placement

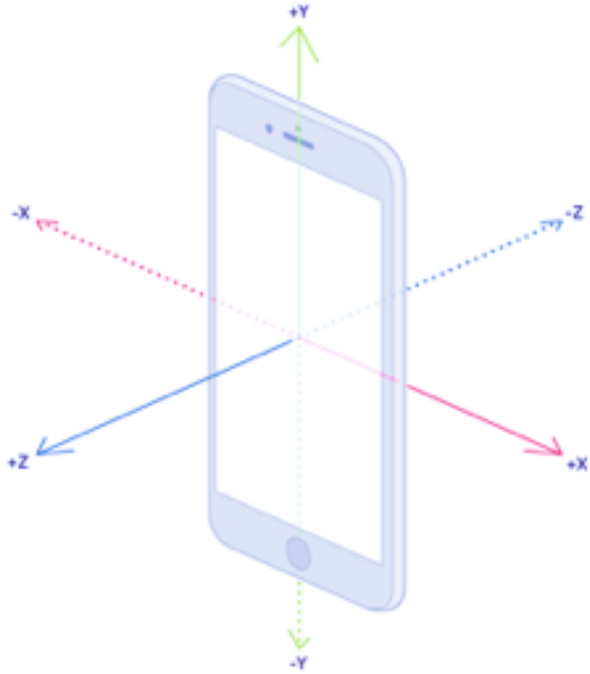

Figure 28: Vehicle's axis

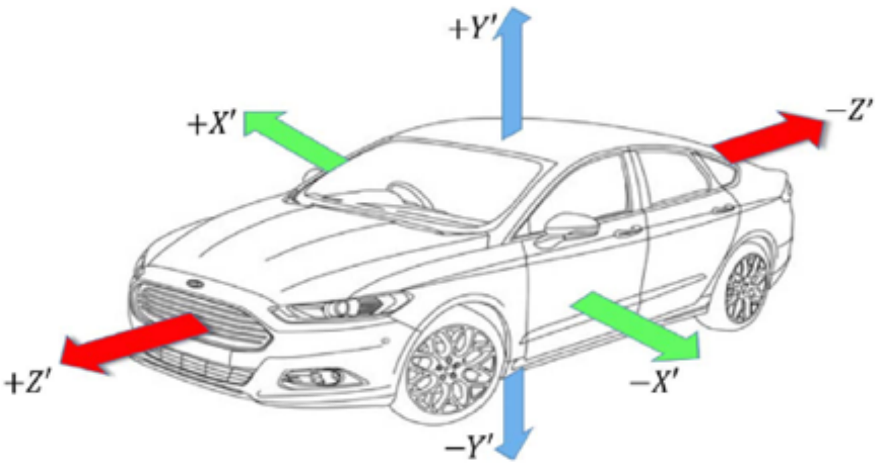

Figure 29: Cartesian coordinate axes of iOS smartphone accelerometer

4.2.3 Speed Dependency

Inoue & Jiang (2017) discussed the issue of motion blur when recording an object from a moving vehicle (Inoue & Jiang, 2017). The degree of motion blur is related to camera exposure and the speed of the target. This study is using images of road potholes that are stationary. The only motion blur could have been caused due to the speed and vibration of the vehicle. The motion blur's degree can be reduced by keeping the exposure time low and the shutter speed fast. This will reduce the brightness of the videos and consequently the road pothole detection rate.

In (Douangphachanh & Oneyama, 2013),the authors discuss how the average speed of a vehicle plays an important role when measuring road roughness. The speed of the vehicle influences the road anomaly detection rate when using accelerometer data. The amplitude of the signal captured by a smartphone accelerometer when a vehicle passes over a pothole depends on the vehicle's speed. For this study, data was collected from all kinds of roads, as mentioned in Table 2. However, most potholes (95%) were recorded on B-roads and in town.

The potholes on the motorway and A-roads were not included because, at the speed of 60 miles/hour, a vehicle will cover over 26 meters per second. Practically it will not be common to find such a big pothole. Also, the current sampling rate is restricted to 100 Hz, which will not be sufficient to capture a pattern in accelerometer data. The smartphone (iPhone Xs Max, iOS) used in this study can record 4K video at 60 frames per second. However, GPS sampling is restricted to 1 Hz. This study aims to detect road potholes using a dashboard camera and tag them with their GPS location. So, pothole data used in this thesis was collected while driving the vehicle at 30 miles per hour. However, normal road surface images were taken from motorways and A roads too.

4.3 Data Pre-Processing

The pre-processing of data is a prerequisite for obtaining good results with high accuracy when developing Machine Learning models. The data pre-processing for the accelerometer data was broken down into three stages:

- Resampling,
- Labelling and,
- Augmentation.

This study aimed to use Convolution Neural Networks to process raw data as they are derived from the smartphone without applying many data processing methods. (Chao, et al., 2020) applied many threshold filters to produce a clean dataset. The images and videos frames obtained from the smartphone camera had different sizes depending upon their makeup and model. The thesis used 128x128 size images in the convolution neural network. The following steps were taken to prepare the training dataset.

4.3.1 Sensory Data

Figure 30 shows a sample of the sensory data (accelerometer data) on the X, Y, and Z axes collected using an iOS smartphone app. The highted blue colour box shows that the sampling rate was set to 100 Hz.

	Timestamp	accelX	accelY	accelZ	Lat	Long
0	25-Dec-2020 13:23:37.300	0.0448	-0.9809	0.0450	52.4015	-0.7424
1	25-Dec-2020 13:23:37.309	0.0754	-0.9740	0.0327	52.4015	-0.7424
2	25-Dec-2020 13:23:37.319	0.1065	-0.9577	0.0020	52.4015	-0.7424
3	25-Dec-2020 13:23:37.330	0.1048	-0.9655	0.0026	52.4015	-0.7424
4	25-Dec-2020 13:23:37.340	0.0679	-0.9847	0.0041	52.4015	-0.7424
5	25-Dec-2020 13:23:37.350	0.0646	-1.0278	0.0624	52.4015	-0.7424
6	25-Dec-2020 13:23:37.360	0.0708	-1.0407	0.0764	52.4015	-0.7424
7	25-Dec-2020 13:23:37.370	0.0699	-1.0420	0.0791	52.4015	-0.7424
8	25-Dec-2020 13:23:37.380	0.0569	-1.0174	0.0507	52.4015	-0.7424
9	25-Dec-2020 13:23:37.390	0.0789	-0.9978	0.0398	52.4015	-0.7424

Figure 30: Sensor data sample

4.3.1.1 Resampling

It was observed that the smartphone was not able to sample the accelerometer data at the fixed frequency uniformly. The sampling rate in the iOS application's accelerometer was set at 100Hz, but it was noted that the accelerometer was sampled in the range of 70-100 Hz. The 1D CNN takes fixed-length data and, hence, to have a consistent sampling rate at 100 Hz, the data with lower sampling rates would have to be deleted. However, for this study, the data were resampled at 100 Hz, and missing values were filled by interpolating the data uniformly. The inbuilt library of SciPy (spline interpolation) (SciPy, 2020) was used for this purpose.

4.3.1.2 Labelling

In this section, a 1D-CNN has been used to detect potholes. Accurate labelling of the data is an essential aspect of supervised learning. The performance of a supervised learning model depends on data and data labelling (Graham & Drobnjak, 2018). The images with potholes were identified manually. Later, the sensor data was labelled using a software script to match the date and time from the photos (Figure 31) to date and time of the sensor data (Figure 32 and Figure 33). The GPS location and timestamp were the information used to tag the dataset and produce tuples of 100 samples (sampling rate). The data were recorded at the sampling rate of 100 Hz on all three axes (X, Y, Z). A one-second data sample has 100x3 timestamps. Figure 34 shows the accelerometer reading on the Y axis. This reading represents a significant movement compared to the readings on X and Z axes. Figure 34 shows a 4-second data sample with pothole detection in the 3rd sec (200-300). It can be noticed that the accelerometer reading on the Y axis had a significant dip, stamps 200-300 when the pothole was detected. The whole 1-sec window was labelled as a pothole.

Image_Name	Day	Time	Lat	Long
IMG_6729.JPG	2020:12:25	13:24:37	52.4029	-0.7426
IMG_6732.JPG	2020:12:25	13:24:40	52.4029	-0.7427
IMG_6733.JPG	2020:12:25	13:24:41	52.4029	-0.7428
IMG_6734.JPG	2020:12:25	13:24:42	52.4029	-0.7428
IMG_6735.JPG	2020:12:25	13:24:43	52.4029	-0.7429
IMG_6736.JPG	2020:12:25	13:24:44	52.4029	-0.7429
IMG_6737.JPG	2020:12:25	13:24:45	52.4029	-0.7429
IMG_6738.JPG	2020:12:25	13:24:46	52.4029	-0.7430
IMG_6739.JPG	2020:12:25	13:24:47	52.4030	-0.7430
IMG_6740.JPG	2020:12:25	13:24:48	52.4030	-0.7430
IMG_6741.JPG	2020:12:25	13:24:49	52.4030	-0.7431

Figure 31: A sample of imagery data

Time	accelX	accelY	accelZ	Damage
11:36:59	0.094200	-1.016600	0.052500	0.0
11:36:59	0.090200	-1.005200	0.056900	0.0
11:36:59	0.074000	-0.985400	0.069100	0.0
11:36:59	0.080629	-0.985738	0.063325	0.0
11:36:59	0.066400	-0.978200	0.074900	0.0
...	...	...	...	...
12:39:53	-0.028997	-0.947385	0.082865	0.0

Figure 32: Sensory data - road surface no damage (normal)

Time	accelX	accelY	accelZ	Damage
11:40:58	0.084921	-1.136693	0.340660	1.0
11:40:58	0.085069	-1.136522	0.340040	1.0
11:40:58	0.085218	-1.136350	0.339421	1.0
11:40:58	0.085367	-1.136178	0.338801	1.0
11:40:58	0.085515	-1.136006	0.338182	1.0
...	...	...	...	...
12:37:43	-0.061500	-1.155600	0.260000	1.0
12:37:43	-0.252400	-1.400900	0.503100	1.0
12:37:43	-0.431000	-1.323200	0.485300	1.0
12:37:43	-0.330000	-1.264800	0.477300	1.0
12:37:43	-0.029700	-1.087800	-0.004900	1.0

Figure 33: Sensory data - road surface damage (pothole)

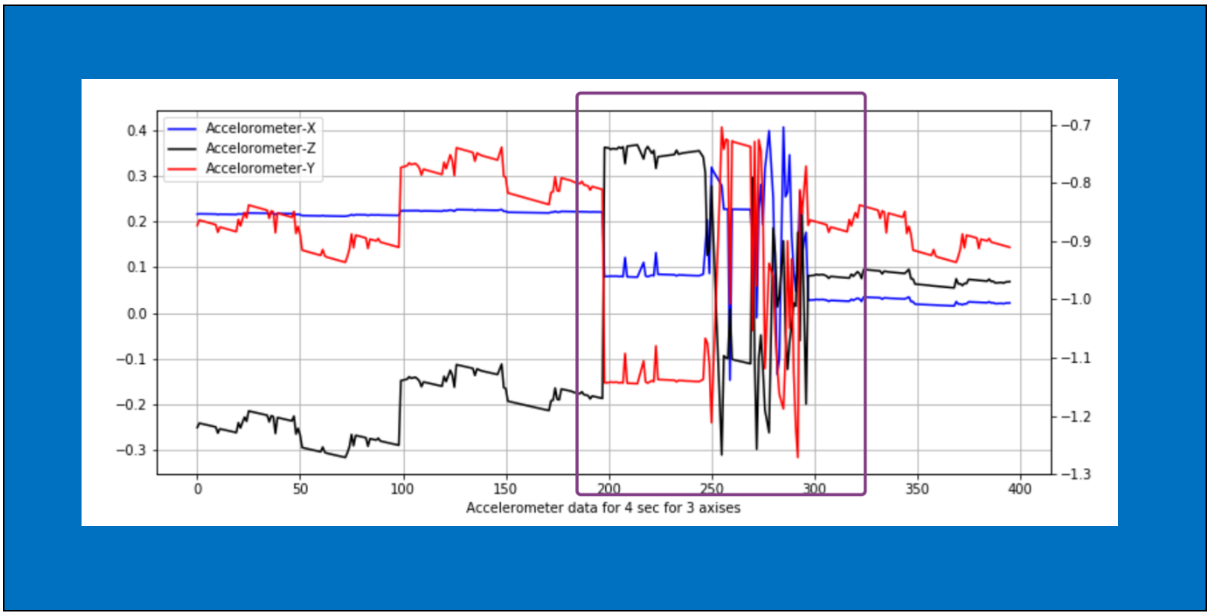

Figure 34: Labelling of a pothole on accelerometer data

4.3.1.3 Data Augmentation

To increase the size of the dataset, data augmentation methods such as permutation and scaling were applied. The standard scaling factors +/-5% were used to emulate if data was collected using different vehicles. The permutation on the dataset was applied to increase the number of potholes.

4.3.2 Imagery Data

The images and videos frames obtained from the smartphone camera had different sizes depending upon their makeup and model. The proposed research has used 128x128 size images as an input to the two-dimensional convolution neural network. The following steps were taken to prepare the dataset.

4.3.2.1 Image Size Correction

The images for the dataset were obtained from more than one sources. Hence, sizes of the images were not same. The first step was to make these images to the same size and remove images that were not fit for the experiment. The first few experiments were conducted using images with a size of 256x256. Later the image size was changed to 128x128 to experiments with the models and to analyse the results. The effect of changes in image size has been discussed in the result section.

Table 5: First data set

Categories	Number of Images
DM00	2814
DM01	4082
DM02	5366

Table 6: Number of images according to size groups

Group	Size	Number of Images
G1	0-10 KB	1883
G2	10-50 KB	1391
G3	50-100KB	411
G4	100-1000 KB	913
G5	1000-2000 KB	37
G6	2000-3000 KB	7
G7	3000-5000 KB	4
G8	5000+ KB	158

Table 5 shows the number of images in each dataset before any modification. Table 6 shows how these images were distributed according to their sizes.

The images which were bigger than the required size (256x256) were split into two images and later, all images were inspected manually to remove if they were not suitable for the experiments. Once all images were changed to 256x256 size, the next step was to label these images accurately, as discussed in the next section. Table 3 shows the data set's details with image size 256x256 and accurate labelling, which is discussed in the next section.

4.3.2.2 Labelling

Accurate labelling of the data is an essential aspect of supervised learning. The accuracy of detection depends on how accurately data is labelled (Doungphachanh & Oneyama, 2013). The images were inspected manually to remove images with high motion blur or unwanted features. The images with potholes were identified and labelled into three classes: no pothole (DM00), pothole (DM01) as seen from the dashboard camera, and bigger potholes (DM02) which were captured using a handheld camera.

Figure 35 and Figure 36 show the images of normal surface and potholes, respectively, captured from the dashboard camera, and Figure 37 shows the images of potholes captured using a handheld device.

Figure 35 : Normal road surface (dm00)

Figure 36: Damaged road surface (DM01)

Figure 37 : Damaged road surface (DM02)

After labelling the images accurately, few image processing methods were used to perform augmentation on these images to increase the size of the dataset. The augmentation processes are discussed in detail in the next section.

4.3.2.3 Data Augmentation

The dataset was prepared carefully to include images from various weather conditions, such as dry road and pothole filled with water, and different time of the day to accommodate brightness variation. However, due to the limitation of time and geographical reach, data augmentation was done on the images to increase the number of images in the dataset. The augmentation techniques which are described in the next few sections were applied on the imagery dataset.

4.3.2.4 Horizontal and Vertical Flip

The image processing library of NumPy was used to flip the images, as shown in Figure 38 to pictures in Figure 39

Figure 38: A pothole image to be augmented

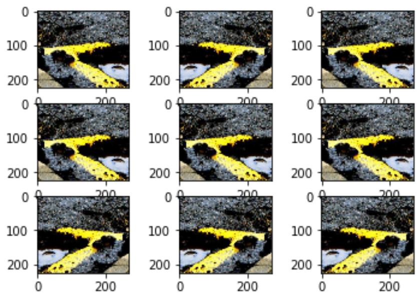

Figure 39: Pothole images after horizontal and vertical flip

4.3.2.5 Random Brightness Transformation

The brightness of the images in the dataset was adjusted in the range of 0.2 to 1 to produce more images. For example, the initial pothole image shown in figure 40 was used to produce the image shown in Figure 41. The brightness of the images was transformed to replicate images from various times when brightness is different.

figure 40: A pothole image to be augmented

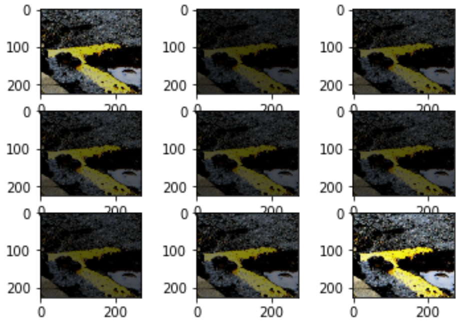

Figure 41: Brightness adjusted images

4.3.2.6 Gaussian Blur and Canny Edge Detector

A canny edge detector is used to extract edges in an image when selecting the area of interest. Before applying the Canny edge detector method, a Gaussian blur was applied to smoothen the images. figure 42(a) shows a sample of an image from the smartphone and (b) shows the same image after applying Gaussian blur to reduce noise. Figure 43(a) shows the same image (figure 42 b) with a Canny edge detector, and Figure 43(b) shows an image with the area of interest, which is within the lane in which the vehicle is moving. The thesis wants to record potholes on-road only, not on the footpath. In Figure 43(b), the black part is the mask to make the area of interest between two white lines clear.

figure 42:(a): Original image, (b): Image after gaussian filter

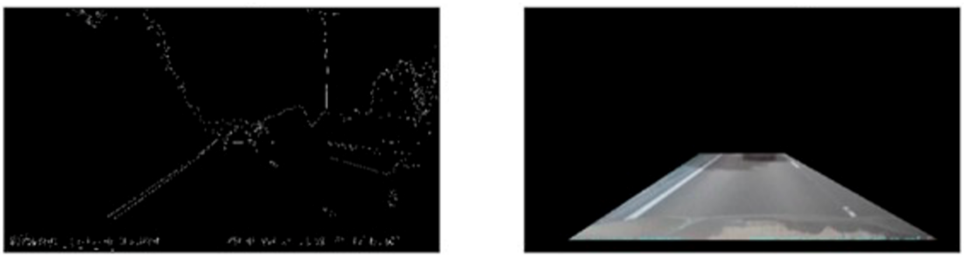

Figure 43:(a): Image with canny edge detector, (b) Area of interest within the canny edge (road lane)

After completing the augmentation processes, the dataset has 2810 images in dataset DM00, 4074 images in DM01 and 3694 images in DM02.

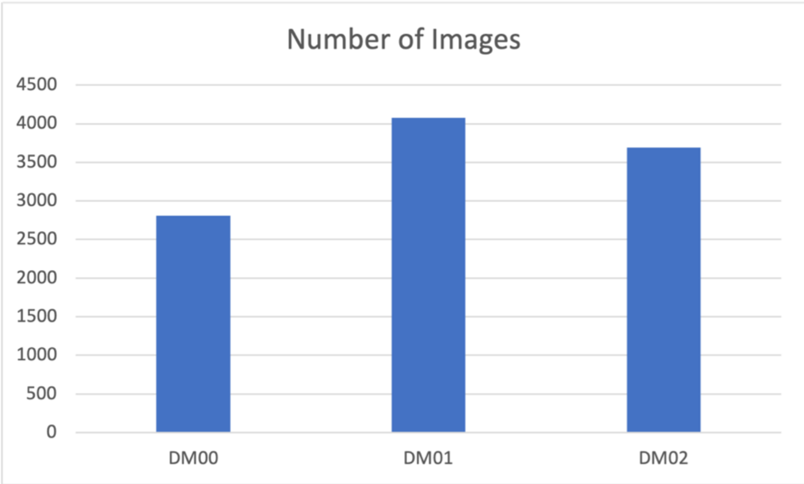

Figure 44: Number of images in each damage class

4.4 Chapter Summary

In this chapter a number of factors were discussed which can affect the identification and classification performance of pothole detection. Briefly, these factors include position of smart phone, orientation, speeding dependency and motion blur. The data augmentation techniques have also been discussed in this chapter which shows how the data was extended in order to run Machine Learning experiments effectively. The data augmentation technique was applied to imagery data set. The data pre-processing techniques and various filters are discussed and applied. These filters included Random brightness transformation and Gaussian blur and canny edge detector.

The next chapter will discuss three Deep Leaning models. More precisely, firstly a two-dimensional Convolution Neural Network will be discussed. Secondly, a one-dimension Neural Networks model will be discussed. Finally, a Fusion model based on Convolution Neural Networks will be discussed. The next chapter will also present the results generated from these models.

CHAPTER 5

ROAD POTHOLE DETECTION USING DEEP LEARNING

5.1 Introduction

As discussed in CHAPTER 2, there are researchers who have used sensory data and Machine Learning such as Support Vector Machine (SVM) and Convolution Neural Networks methods to classify road surface. However, no paper has used sensory data and Convolution Neural Networks to detect road potholes to the best of the author's knowledge. Also, many papers discussed image classification using Machine Learning methods. There are few papers which have used Convolution Neural Networks to classify road surface and detect potholes. However, no paper has used imagery data captured from a standard smartphone (used as a dashboard camera) and Convolution Neural Networks to classify road surface to the best of the author's knowledge.

This research aims to experiment with sensory or accelerometer data and one-dimensional Convolution Neural Networks to develop a model with the optimal performance parameters. Moreover, this research aims to use imagery data to experiment with a two-dimension Convolution Neural Networks to develop a suitable network architecture and network parameters. Following the development of the two individual models for sensory data and imagery data. A novel fusion model is developed based on ID-CNN and 2D-CNN models. The fusion model takes sensory data and imagery data as two inputs and produces an output.

The rest of the chapter is organised as follows. Section 5.2 discusses the experiments of imagery data with a two-dimension Convolution Neural Networks. Section 5.3 presents the experiments on sensory data with one-dimensional Convolution Neural Networks. Finally, Section 5.4 presents the new fusion model containing one dimension Convolution Neural Networks and two-dimensional Convolution Neural Network.

5.2 Deep Learning Model to Detect Road Pothole using Imagery Data

This study has employed a Convolution Neural Networks model with multiple layers of non-linear transformation ranging from three to five to obtain the classification results. Figure 45 shows the DL network with five hidden layers. Kernel size 5x5, stride 2x2, and ReLU (Rectified Linear Unit), activation function was utilized on the hidden layers and on the output layer, the SoftMax activation function was used. The loss function was categorical cross-entropy, and Adam optimizer was used. The SoftMax function transforms a fully connected layer into the probability distribution for classification among the classes under investigation. The Categorical Cross-Entropy Loss for the number of classes C is defined as

$$J(X_i Y_i, \theta) = -\sum_{j=1}^{C} y_{ij} * \log(p_{ij}). \text{ --} \textit{Eq.3}$$

Where Y_i is the hot encoded target vector. y_{ij} will be one if i_{th} element is in class j; otherwise, it will be zero. p_{ij} is the probability for the same.

SoftMax function defined by $f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}}$ is used to find the probability p_{ij}

$$\theta = \theta - \eta . \nabla_\theta J(x^i, y^i, \theta) \text{--Eq.4}$$

Where $\boxed{\boldsymbol{\theta}}$ is a weight parameter, the training aims to minimize loss (Eq.3) and get optimal weight parameter $\boxed{\boldsymbol{\theta}}$.

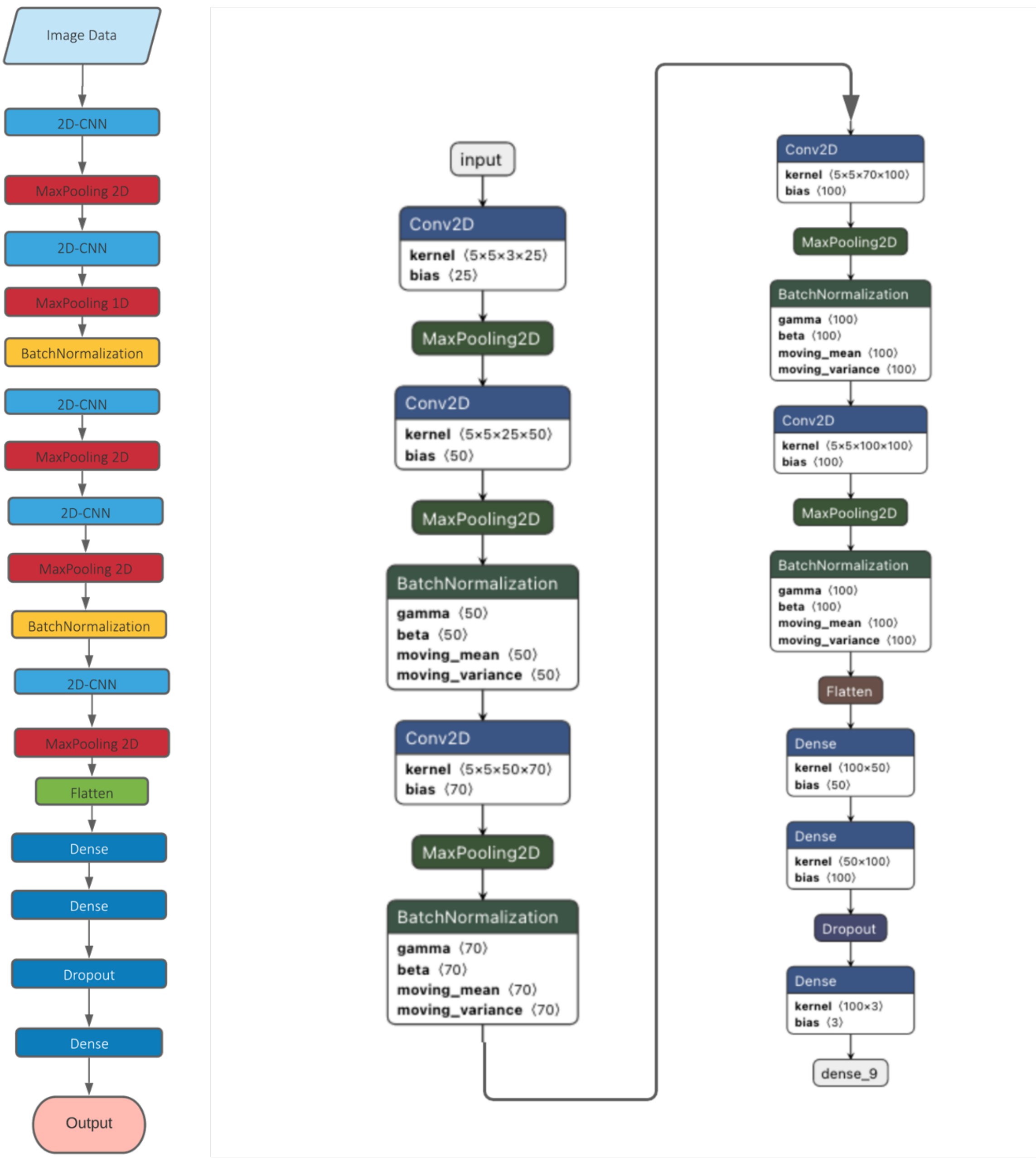

Figure 45: Convolution neural network model with five layers

Table 7 shows the simplified form of the algorithms and sequences of activities which were used to train imagery data and sensory data for Convolution Neural Networks.

Table 7:Algorithm to train a convolution neural networks

Algorithm: To train a Convolution Neural Networks

Input: Labelled:{ X,Y}

Output: Optimal θ^{*};

- Initial $\boxed{\theta}$, epoch = 0, learning rate α
- repeat
- Sampling labelled data batch $\{x_i, y_i\}$ from $\{X, Y\}$
- Performing forward propagation of the network
- Compute loss by Eq.3
- Compute adaptive gradient by SDG (Eq:2)
- Update parameter $\boxed{\theta \leftarrow \theta - \alpha \frac{\partial L}{\partial \theta}}$ $\theta \leftarrow \theta - \alpha \frac{\partial J}{\partial \theta}$
- epoch = epoch + 1
- until (epoch > Epochs)

Figure 46 shows the steps involved in identifying an image. Table 8 describes Algorithm to identify an image.

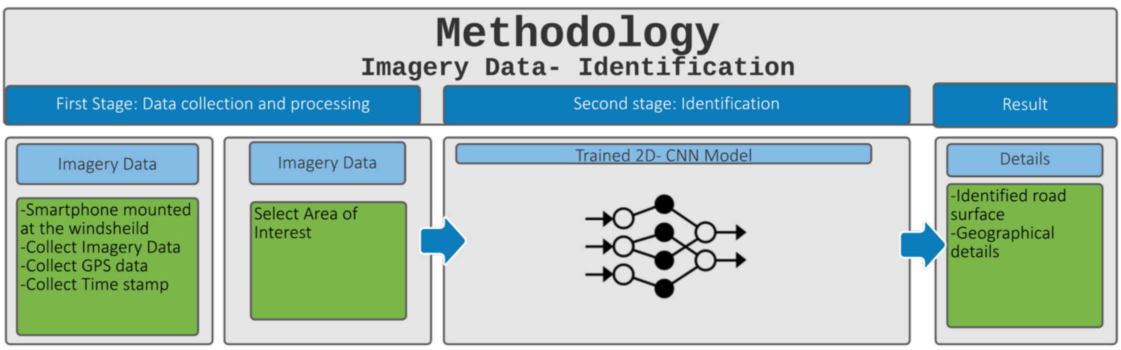

Figure 46:Steps to identify an image

Table 8:Algorithm to identify road pothole using imagery data

Algorithm: Road pothole Identification using sensory and convolution neural network

Input: Images with GPS location

Output: Pothole category (DM00, DM01, DM02), GPS location

1. Capture Images using a smartphone camera
2. Upload images on a cloud server
3. Download images from the server to a computer
4. Select the area of interest (4.2.6.3)
5. Feed images from step 4 into the Identifier model
6. The categorical output from the model
7. Tag the classified images with GPS details.
8. Update the database
9. Send notification to other vehicles (who have opted for) which are in proximity.

The dataset Table 3 was split into 70% for training, 15% for testing and 15% for validation and then was used to train and test the Convolution Neural Networks model with various combinations of hyperparameters. The study started by examining a model with two layers and two classes: pothole (DM01) and no pothole (DM00). Later a model with five hidden layers and three classes was considered, as mentioned in Table 3 was used. The Rectified Linear Unit (Relu) activation function was used on hidden layers and SoftMax on the output layer. Each training was conducted with batch size 10, the number of epochs 50 and dropout 0.5. The analysis involved a comparison of test accuracy, precision and recall for various models.

The experiment (5.2.1, 5.2.2, 5.2.3, 5.2.4, 5.2.5) were conducted with image size 256x256, batch size 50 and epoch 50. The simulations were executed on a mac with configuration shown in Table 9.

Table 9:Mac desktop configuration

Mac Desktop Configuration	
Processor	3.2 GHz Quad-Core Intel Core i5
Memory	24 GB 1867 MHz DDR3
Graphics	AMD Radeon R9 M390 2 GB

5.2.1 DL Model with 5 Hidden Layers 2 and 2 Classes

It took over 7 hours to run this experiment and the test accuracy of the model was 60.81%. The accuracy was low, so in the following experiment the number of hidden layers in the model was increased to check whether the result will improve.

Number of Layers	2
Number of Classes	2 (DM00, DM01)
Test Accuracy	60.81%
Time	7:42:51

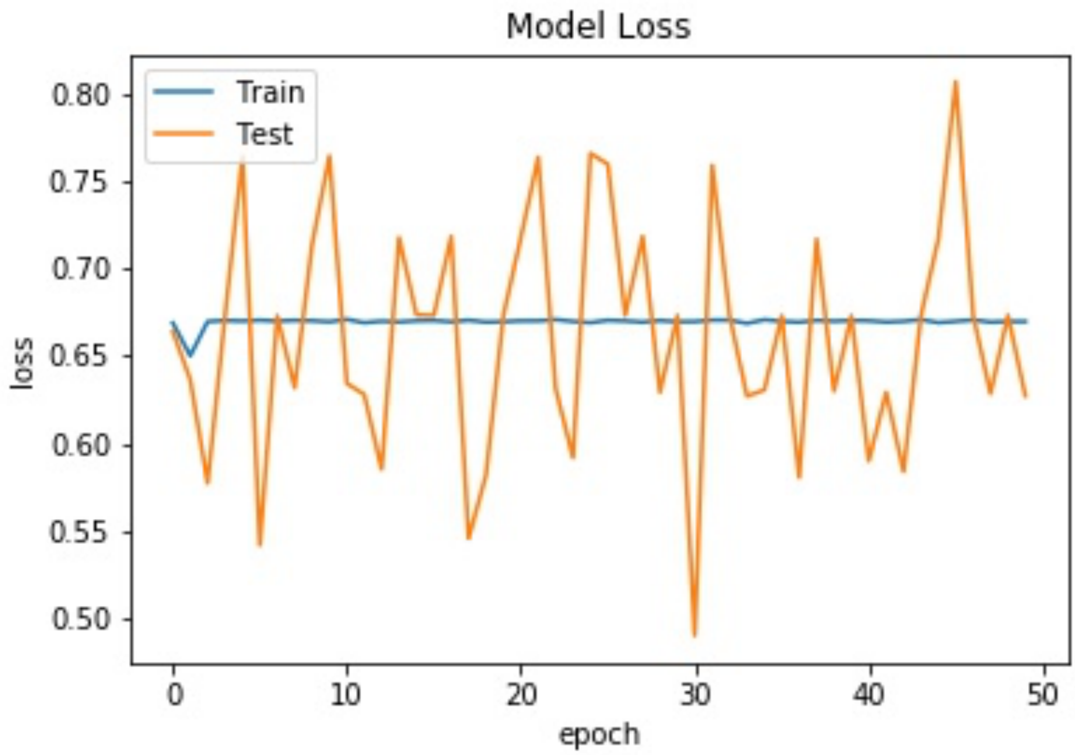

Figure 48 :2D-CNN model loss with 2 HL and 2 classes

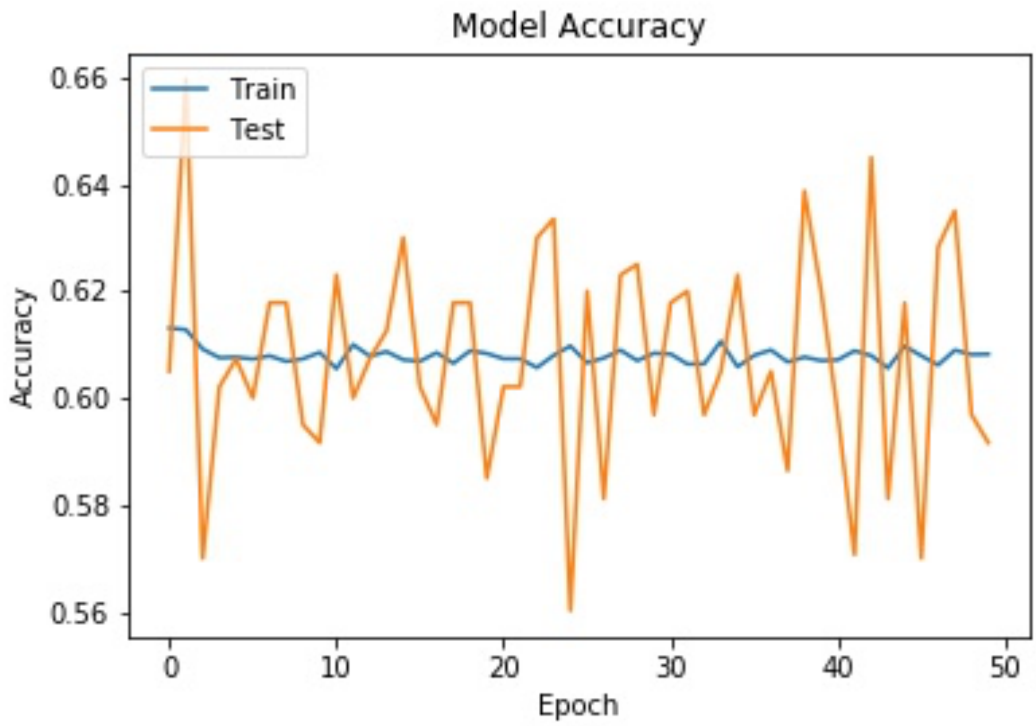

Figure 47: 2D-CNN model accuracy with 2 HL and 2 classes

5.2.2 DL Model with 5 Hidden Layers 5 and 2 Classes

The previous experiment's result was not in the excellent range to the number of hidden layers in the model was increased to check if the accuracy improves.

Number of Layers	5
Number of Classes	2 (DM00, DM01)
Train Accuracy	98.97%
Test Accuracy	97.81%
Time	20:32:12

This experiment took over 20 hours to run on the mac computer described in Table 9 and obtained a maximum test accuracy of 97.81 %. However, as it can be noticed from the graphs in Figure 49 and Figure 50 the accuracy and loss of the model is not consistent.

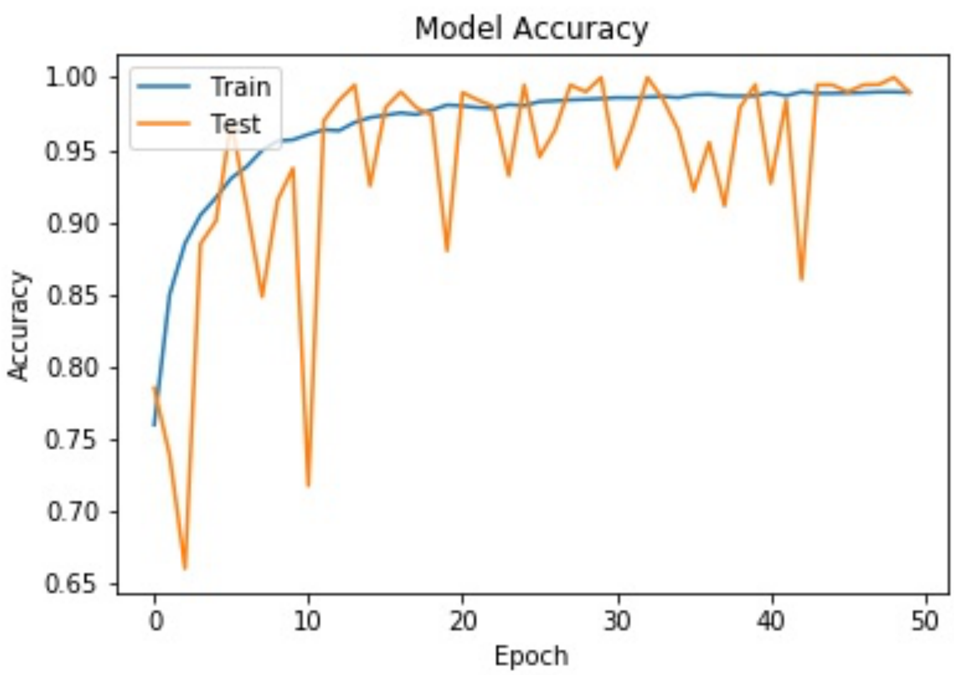

Figure 49:2D-CNN model accuracy with 5 HL and 2 classes

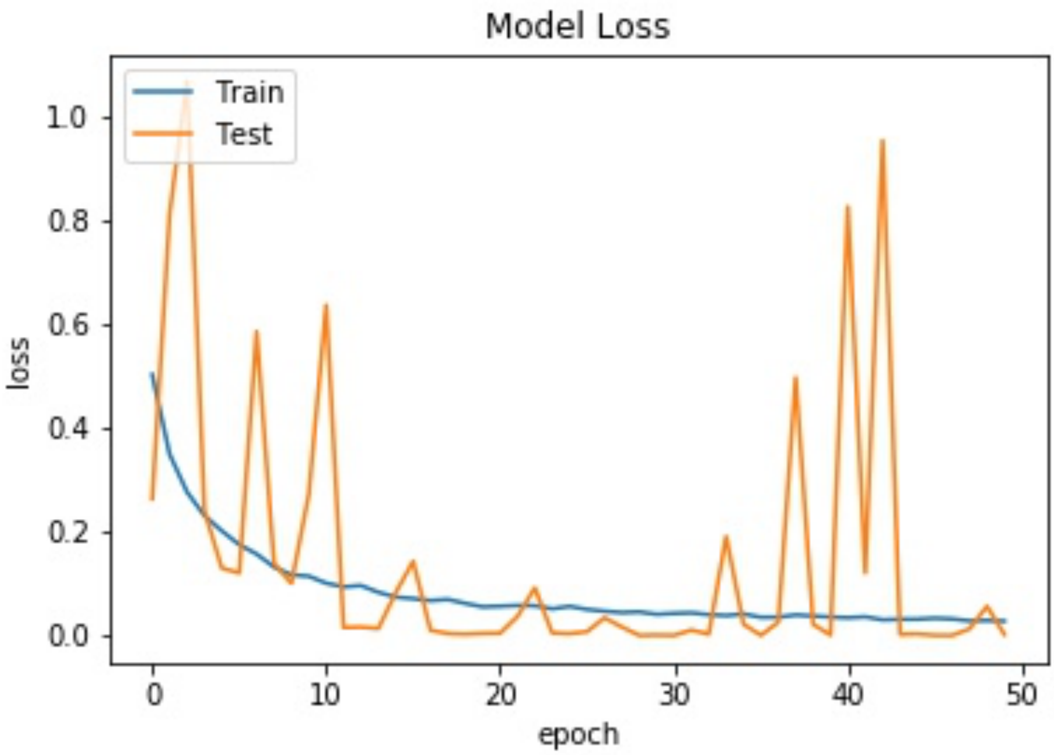

Figure 50: 2D-CNN model loss with 5 HL and 2 classes

5.2.3 DL Model with 5 Hidden Layers and 3 Classes

The following experiment was conducted with 5 hidden layers and three classes; two classes were the same as in the previous experiment and DM02 was also used in this experiment. It took over 25 hours to complete the experiment. The maximum test accuracy was 94.49%.

Number of Layers	5
Number of Variables	3 (DM00, DM01, DM02)
Train Accuracy	96.60%
Test accuracy	94.49%
Time	25:30:10

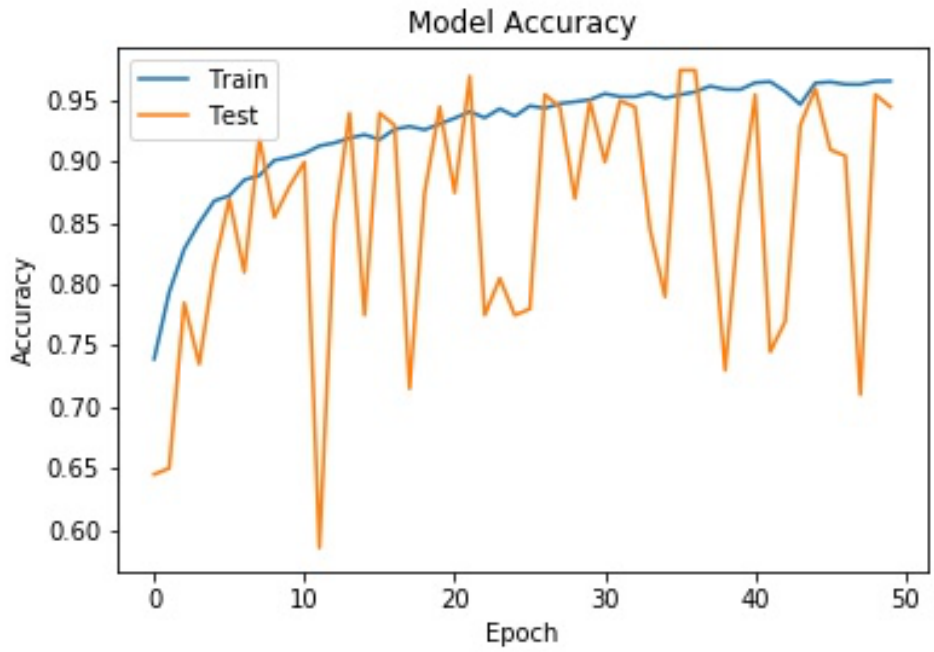

Figure 51: 2D-CNN model accuracy 5 HL and 3 classes

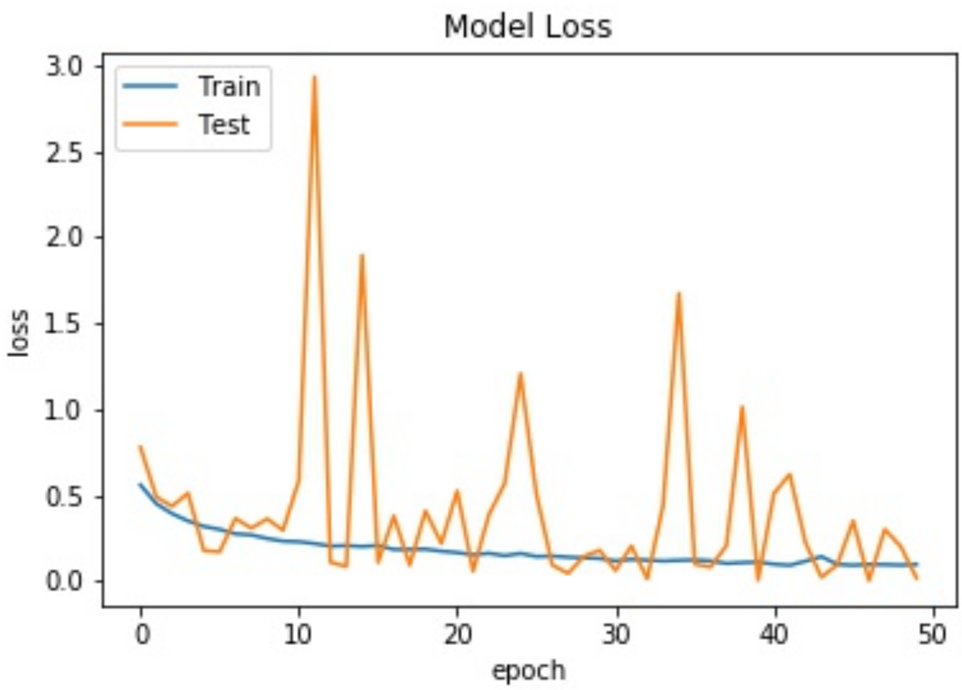

Figure 52: 2D-CNN model loss with 5 HL and 3 classes

The accuracy of this experiment, with three classes and 5 hidden layers, have decreased. The maximum test accuracy is 94.49%. However, the accuracy and loss of the model as it can be seen in Figure 51and Figure 52 are not consistent. So, it was decided to increase the batch size from 10 to 50.

5.2.4 DL Model with 5 Hidden Layers 5 and 3 Classes

This experiment was executed with the same parameters as in the previous experiment, but the batch size was changed to 50 from 10.

Number of Layers	5
Number of Variables	3 (DM00, DM01, DM02)
Train Accuracy	99.81%
Test accuracy	91.27%
Time	3 days, 17:48:38

This model took more than three days to complete on the same computer. However, it did not make the model loss and accuracy smoother. In fact, maximum train accuracy increased to 99.81%, but maximum test accuracy went down to 91.27%. This shows that the model was overfitting.

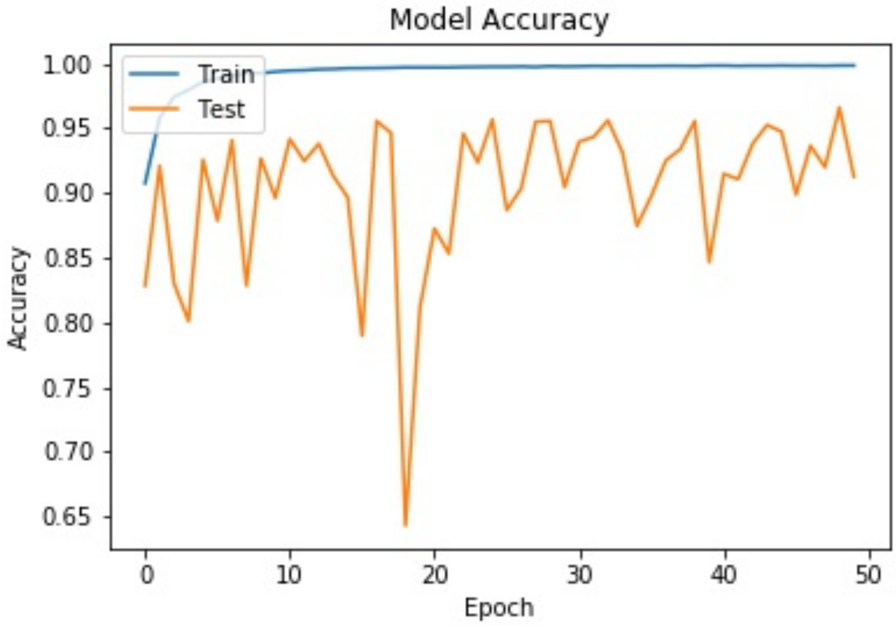

Figure 53:2D-CNN model accuracy with 5 HL and 3 classes

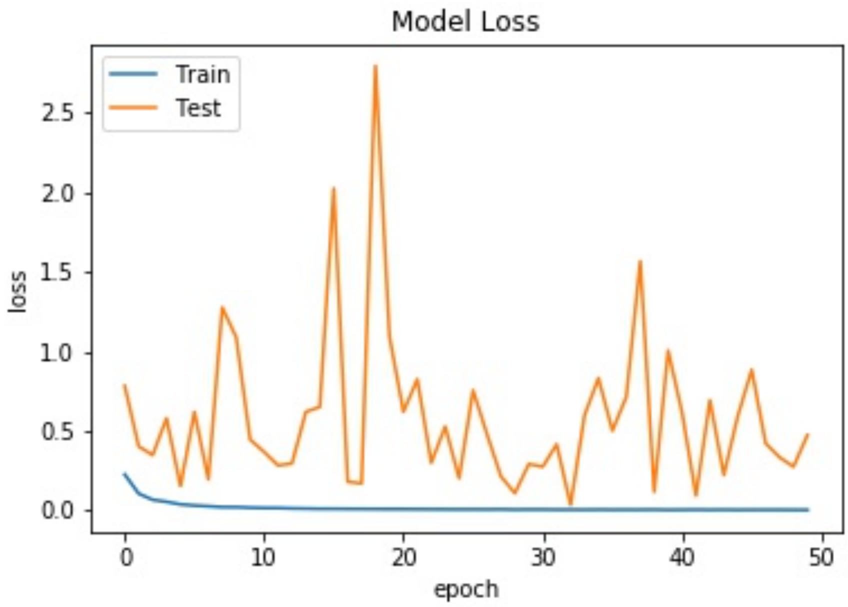

Figure 54: 2D-CNN model loss with 5 HL and 3 classes

The model accuracy Figure 53 and model Figure 54 are not smooth and not consistent. Also, it seems that the model was overfitting. It was decided to change the images' size to check if the model stop overfitting and the accuracy improves.

5.2.5 DL Model with 5 Hidden Layers and 3 Classes

For this experiment, all other model parameters were the same, except the image size, which was changed to 128x128. The experiment took over 16 hours to complete, significantly less than the time needed for the previous experiment. The accuracy also decreased. The difference between train and test accuracy shows that the model was overfitting.

Number of Layers	5
Number of Variables	3 (DM00, DM01, DM02)
Train Accuracy	99.79%
Test accuracy	86.15%
Time	16:52:40

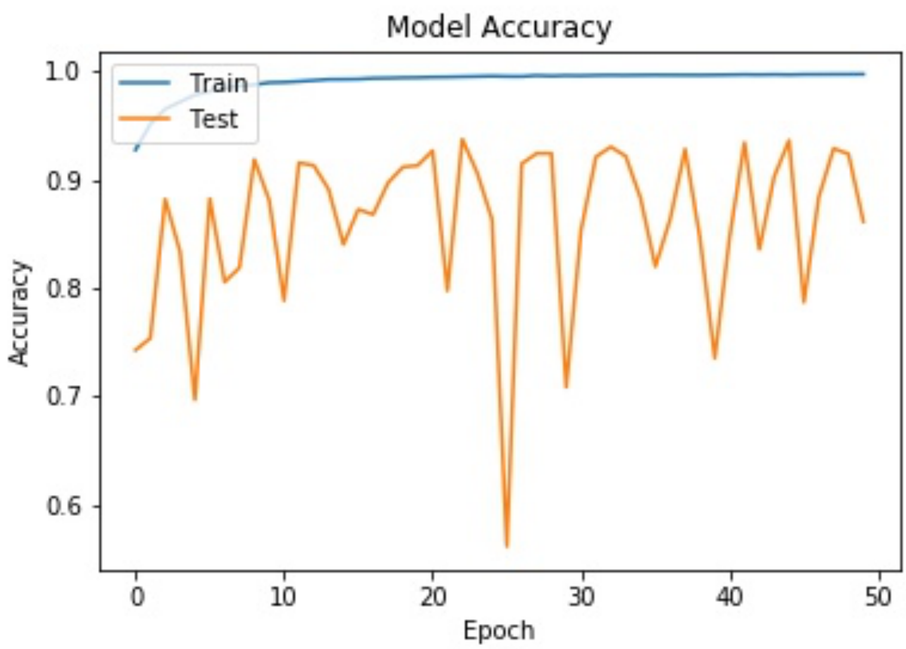

Figure 55: 2D-CNN model accuracy with 5 HL and 3 classes

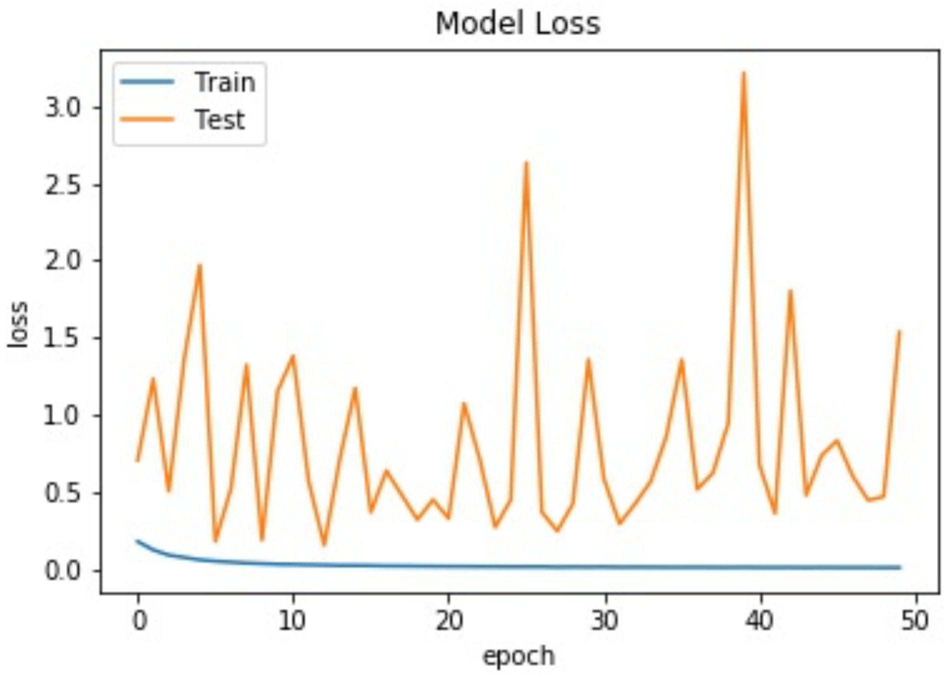

Figure 56: 2D-CNN model loss with 5 HL and 3 classes

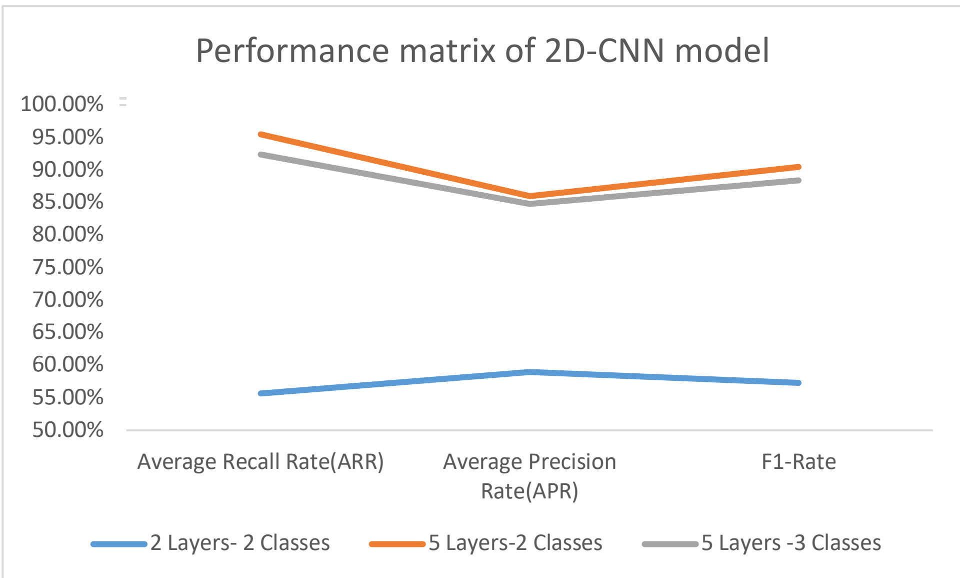

Figure 57: 2D-CNN model performance matrix

5.2.6 Results

The dataset Table 4Table 3 was split into 70% for training,15% for testing and 15% for validation and then was used to train and test the Convolution Neural Networks model with various combinations of hyperparameters. The study started by examining a model with two layers and just two classes: pothole (DM01) and no pothole (DM00). Later a model with five hidden layers and three classes was used. The Rectified Linear Unit (ReLU) activation function was used on hidden layers and SoftMax on the output layer. Each training was conducted with batch size 10, the number of epochs 50 and dropout 0.5. The model's training accuracy with two hidden layers and two classes was in the range of 60.55% and 61.31% and the test accuracy was in the range of 56% and 66% with a median of 60.50%. The accuracy was on the lower side, so another Convolution Neural Networks model with the number of hidden layers five was used to train the same dataset of two variables (DM00 and DM01). The Convolution Neural Networks model's training and test results with hidden layers five are shown in (see Figure 49 and Figure 50).This model's training accuracy was in the range of 75% and 99%, with a median of 98% and the test accuracy was in the range of 66% and 100%, with a median of 97% The same model was trained with three variables (DM00, DM01, DM02) described in Table 3. The model's simulation result with five hidden layers and three classes is shown in Figure 53 and Figure 54. This Convolution Neural Networks model has test accuracy in the range of 64% and 96.59%, with the median at 92.23. The confusion matrix of the model with the various combination is shown in Figure 57. The Convolution Neural Networks model with 5 hidden layers and 3 classes has 84.80% precision, 92.40% recall and 88.44% F1-Score. The result is good considering the dataset's size and variation in the type of images in the dataset. It can be noticed that the model's accuracy with just two classes (DM00, DM01), which were collected from the dashboard camera, is higher than the accuracy for the model with three classes (DM00, DM01, DM02). This could be because the DM02 dataset created with images captured using a handheld camera had many potholes filled with water and were bigger. The accuracy of the 2D-CNN model will increase with the dataset's size and when potholes’ type size, shape and depth have more variation.

5.3 Deep Leaning Method to Detect Road Pothole Using Sensory Data

Convolution Neural Networks (CNN) were predominately designed for images classification. Convolution Neural Networks learn features from two-dimension images and produce classification and forecasting results. As demonstrated by previous research,1D-CNN is able to effectively recognise human activities based on accelerometer data collected using a smartphone. The promising result in the study inspired us to use 1D-CNN to classify pothole. In this study, 1D-CNN is used, as Convolution Neural Networks models can learn well from the raw data. The process requires labelling of images of a pothole. Later a script was used to extract time and GPS information to match the data with accelerometer data to label pothole.

There is no requirement for domain expertise to process the data before feeding them into the model manually. The 1D CNN model can automatically learn the feature of the time series data and produce a good model from the dataset. 1D-CNN models with multiple layers of non-linear transformation ranging from two to seven were used to obtain the classification results. Figure 58 shows the network with two hidden layers. Kernel size 5; dropouts .25 and .5 were used during the study. The loss function binary cross-entropy (Eq.1) and stochastic gradient descent (Eq.2) optimiser were used. The SoftMax, function transform fully connected layer into the probability distribution for binary classification between a pothole and no pothole. 1

$L(x, y, \theta) = -\frac{1}{N}\sum_1^N(-y_p.\log(y_{i-p}^{\wedge}) - y_n.\log(y_{i-n}^{\wedge}))$ ---------------------Eq.1

$\theta = \theta - \eta.\nabla_\theta L(x^i, y^i, \theta)$--Eq.2

Where $\boxed{\boldsymbol{\theta}}$ is a weight parameter, the training aims to minimise loss (Eq.1) and get optimal weight parameter $\boxed{\boldsymbol{\theta}}$. For the sample $\boxed{\boldsymbol{x_i}}$, the predicted negative probability is denoted by $\boxed{y_{i-n}^{\wedge}}$ and positive probability by $\boxed{y_{i-p}^{\wedge}}$

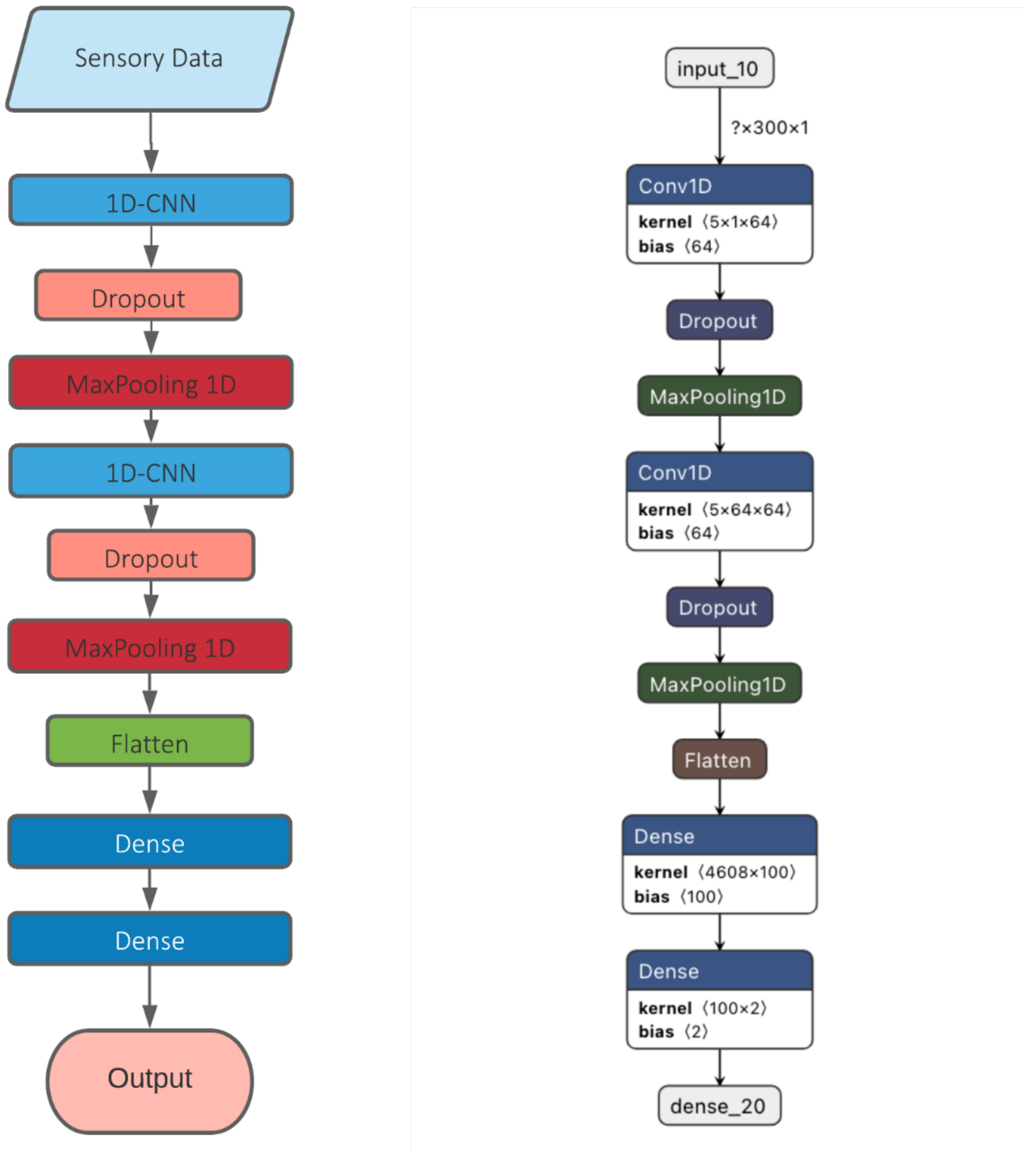

Figure 58: 1D-CNN model with 2 layers

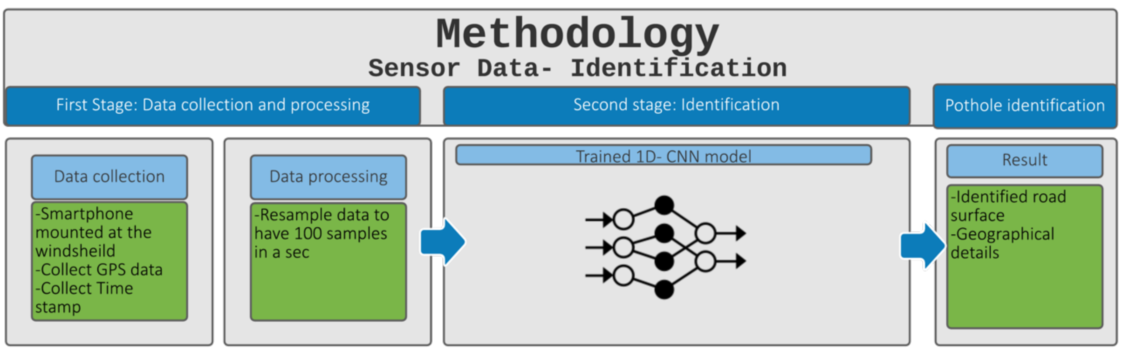

Figure 59:Workflow to classify a sample as a pothole or no pothole

Table 10:Algorithm to identify road pothole using sensory data

Algorithm: Road pothole identification using sensory and one-dimensional convolution neural network

Input: Sensory Data, with GPS

Output: Pothole or No pothole; GPS location

1. Capture sensor data using an accelerometer sensor app of the smartphone
2. Upload data on a cloud server
3. Download data from server to a computer
4. Pre-process data to make that one-sec data sample has 100 timestamps (on all three axis X, Y, Z), if not do resampling
5. Feed data from the previous step to the classifier model
6. If $y_{i-p} > y_{i-n}$ then
7. The fed sample represents a pothole
8. Else
9. No pothole (normal road surface)
10. End if
11. Tag the sample with pothole and GPS details.
12. Upload the classifier on the cloud server
13. Send notification to other vehicles (who have opted for) which are in the proximity.

The experiments with the sensory dataset Table 4 were conducted for the models by using different hidden layers, different kernel size and different dropout. However, each experiment was run for 500 epochs. The details of the experiments are as below.

5.3.1 DL Model with No of Hidden Layers -1 Kernel Size 3 Dropout- 0.25

The first experiment was conducted with a model which has one hidden layer; kernel size was three and dropout was set at .25. The training accuracy was 84.84% and 95.20% and the median was 93%. The test accuracy was in the range of 49.36% to 75.32%, with a median of 91.19%. Figure 60 , Figure 61 and Figure 62 show loss and accuracy and errors of the model. It can be noted that the model obtained a good level train accuracy quite early. However, test accuracy was neither smooth nor consistent. HL = number of hidden layers, KS= Kernel Size and DO= Dropout, will be used to caption the figures.

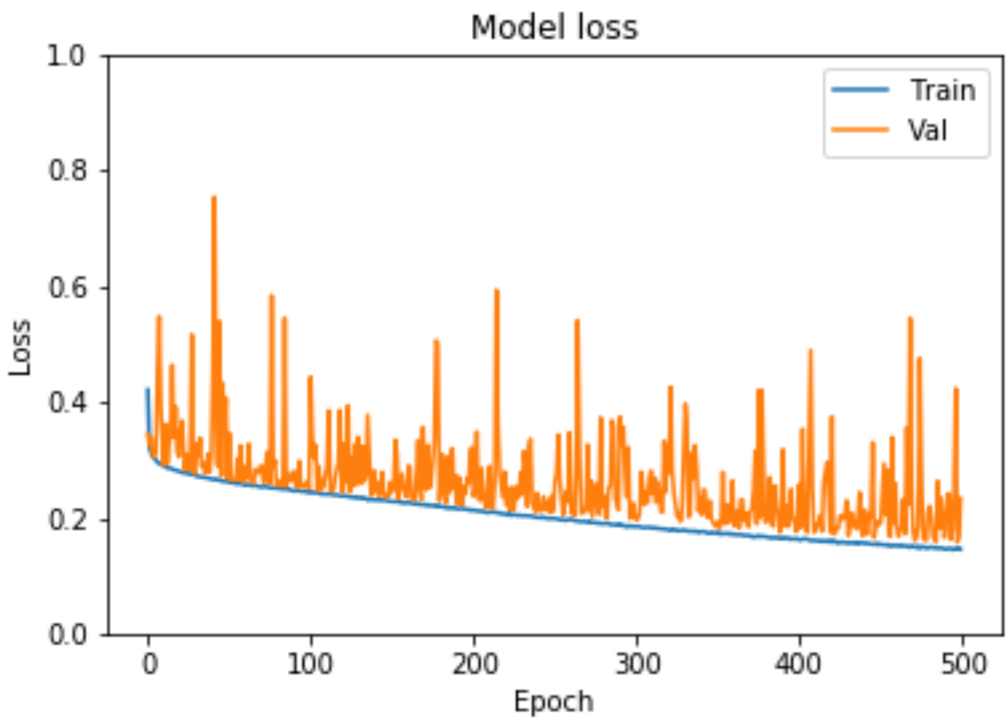

Figure 60: 1D-CNN model loss with HL 1 KS3 dropout .25

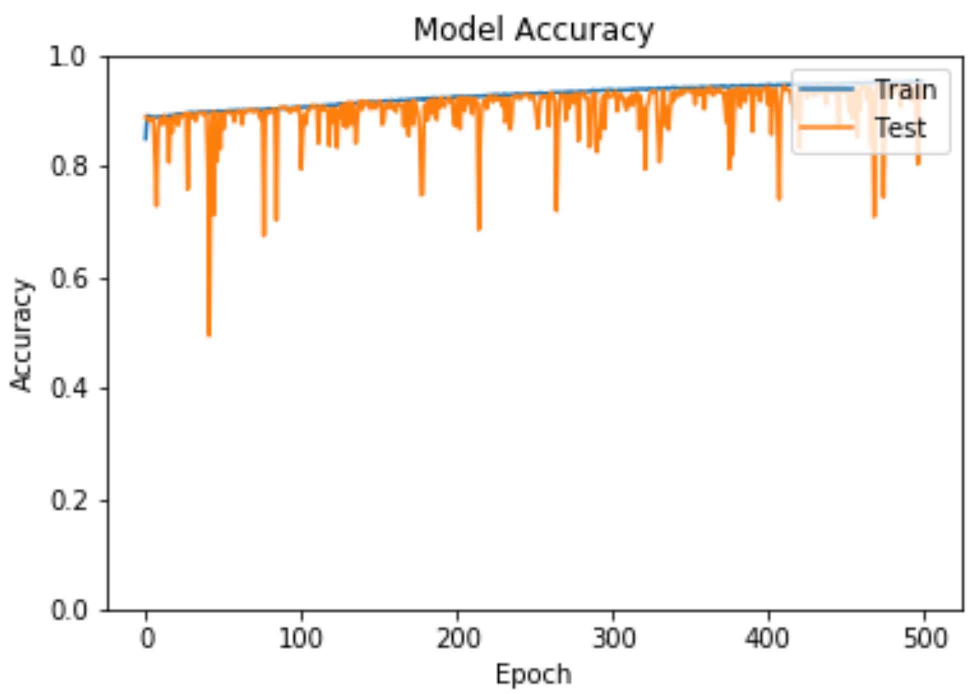

Figure 61: 1D-CNN model accuracy with HL 1 KS3 dropout .25

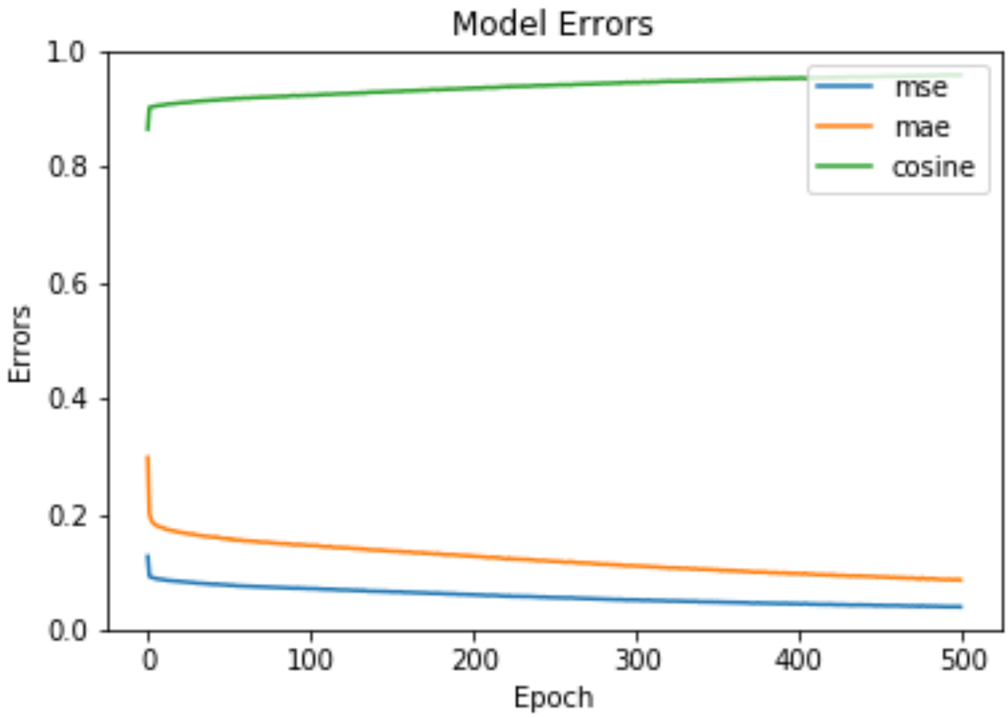

Figure 62:1D-CNN 1D-CNN model errors with HL 1 KS3 dopout .25

5.3.2 DL Model with No of Hidden Layers 1 Kernel Size 5 Dropout 0.25

To address the previous model's issues, a new model with a different kernel size was designed. The same dataset was used to run the experiment. This model's training accuracy was in the range of 81.20% to 95.23% and the median at 92.87%. The test accuracy was in the range of 42.03% to 94.65%, with the median at 91.65%. Figure 63 , Figure 64 and Figure 65 show loss and accuracy and errors of the model. It can be noted that the model obtained the level of good train accuracy quite early. However, the test accuracy of the model was not as good as the previous model.

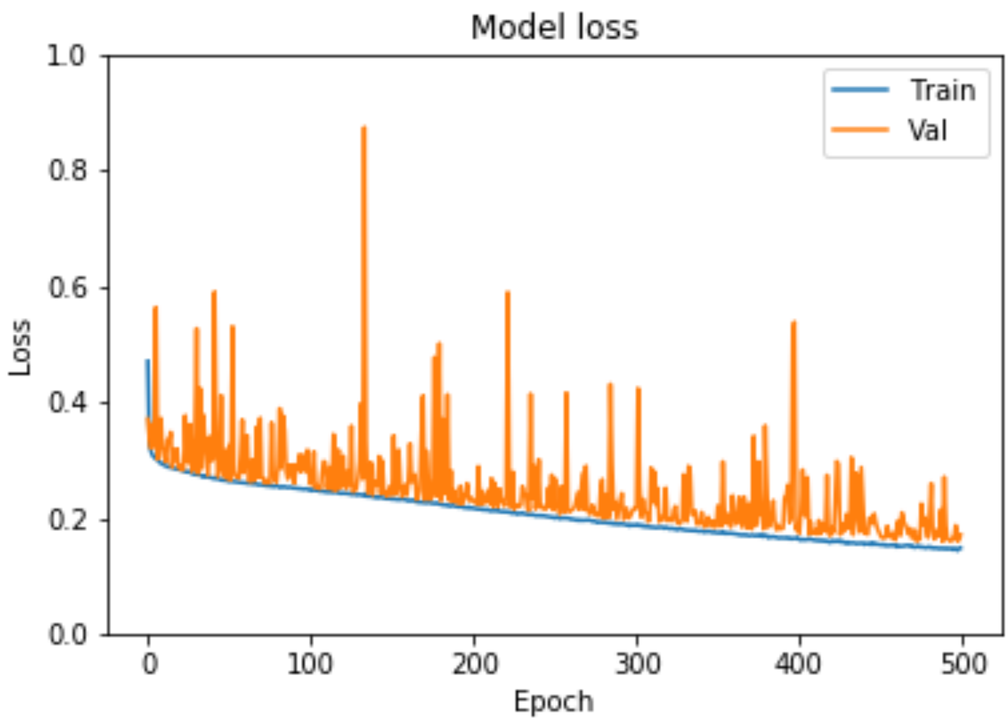

Figure 63: 1D-CNN model loss with HL 1 KS5 dropout .25

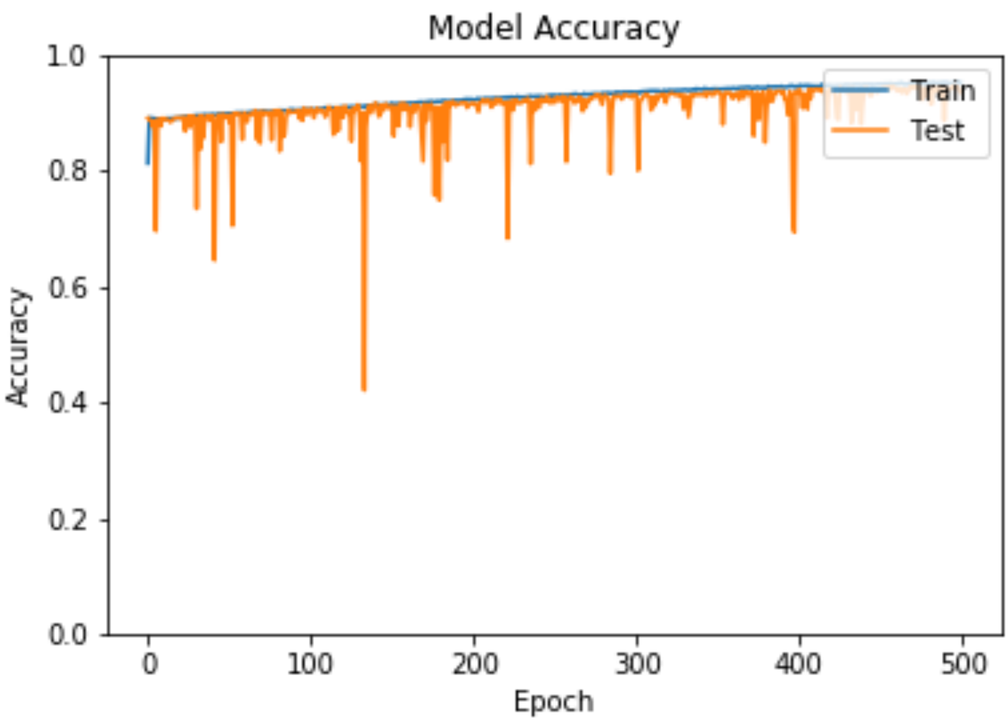

Figure 64: 1D-CNN model accuracy with HL 1 KS5 dropout .25

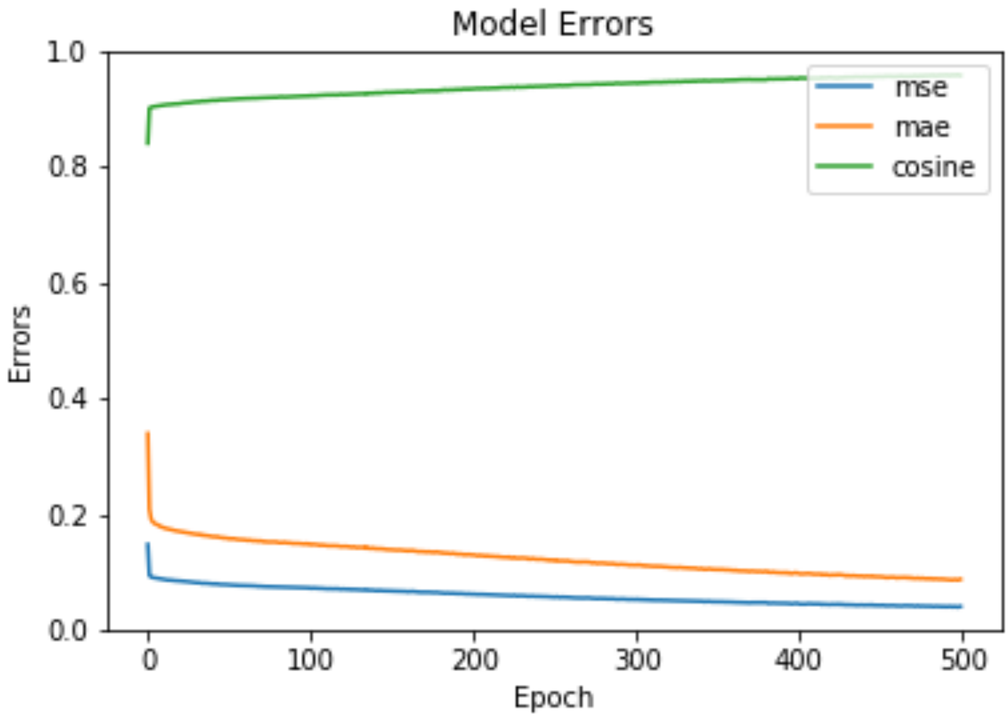

Figure 65:1D-CNN model Errors with HL 1 KS5 dropout .25

5.3.3 DL Model with No of Hidden Layers 2 Kernel Size 3 Dropout 0.25

A new model with hidden layers 2, kernel size 3 and dropout .25 was designed to train at the same dataset. This model's training accuracy was in the range of 66.99% to 95.98% and the median at 66.99%. The test accuracy was in the range of 54.59% to 95.47%, with a median of 92.17%. Figure 66, Figure 67 and Figure 68 show loss, accuracy and errors of the model. It can be noted that the model obtained the level of good train accuracy quite early. However, the test accuracy of this model's performance was not as good as the previous model.

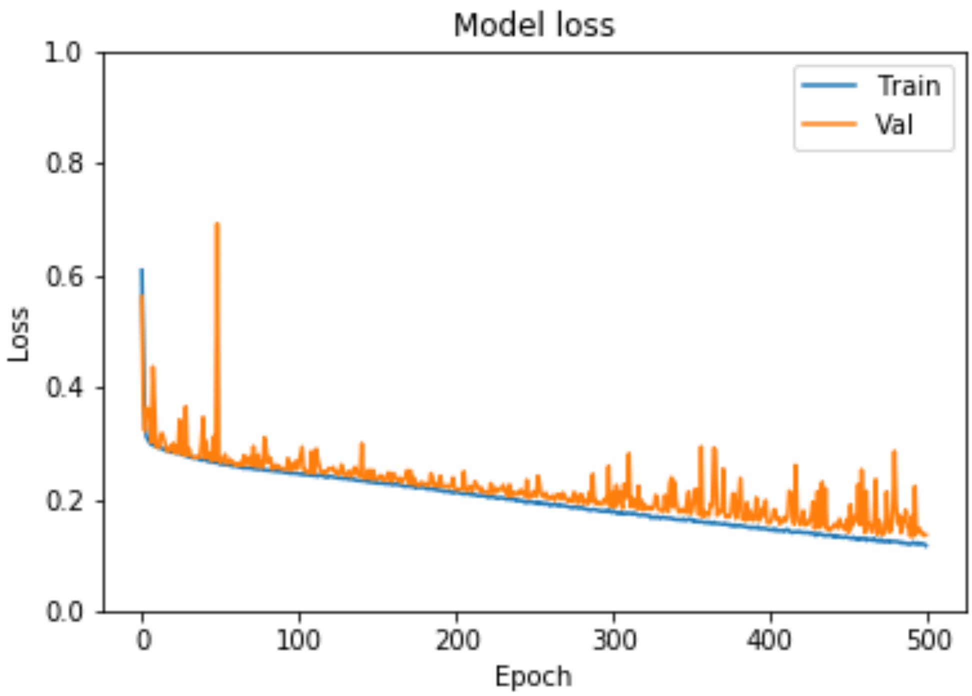

Figure 66: 1D-CNN model loss with HL 2 KS3 dropout .25

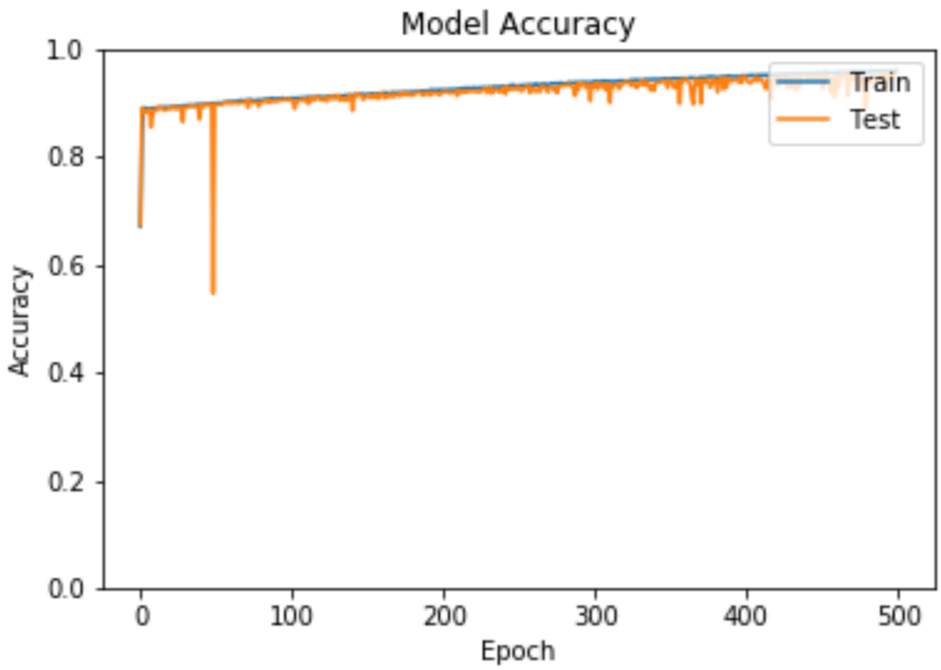

Figure 67: 1D-CNN model accuracy with HL 2 KS3 dropout .25

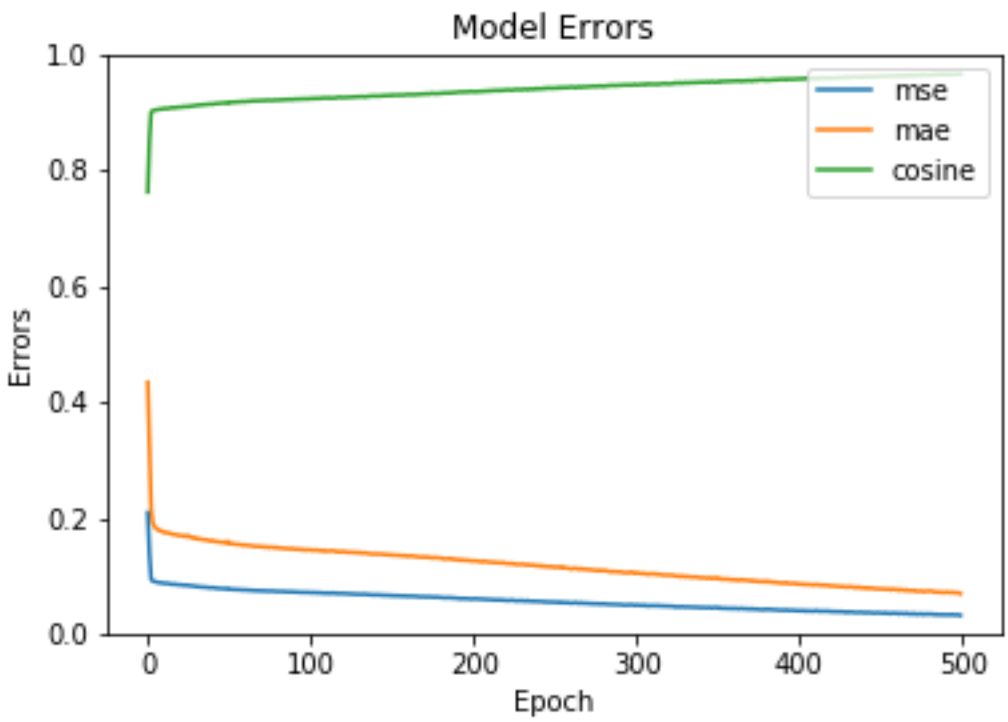

Figure 68:1D-CNN model errors with HL 2 KS3 dropout .25

5.3.4 DL Model with No of Hidden Layers 2 Kernel Size 5 Dropout 0.25

Another experiment with two hidden layers, kernel size 5 and dropout .25, was executed. This model's training accuracy was in the range of 73.38% to 95.91% and the median at 92.98%. The test accuracy was in the range of 82.36% to 95.67%, with a median of 92.03%. Figure 69, Figure 70 and Figure 71 show loss, accuracy and errors of the model. With the increase in Kernel Size, minimum accuracy was improved.

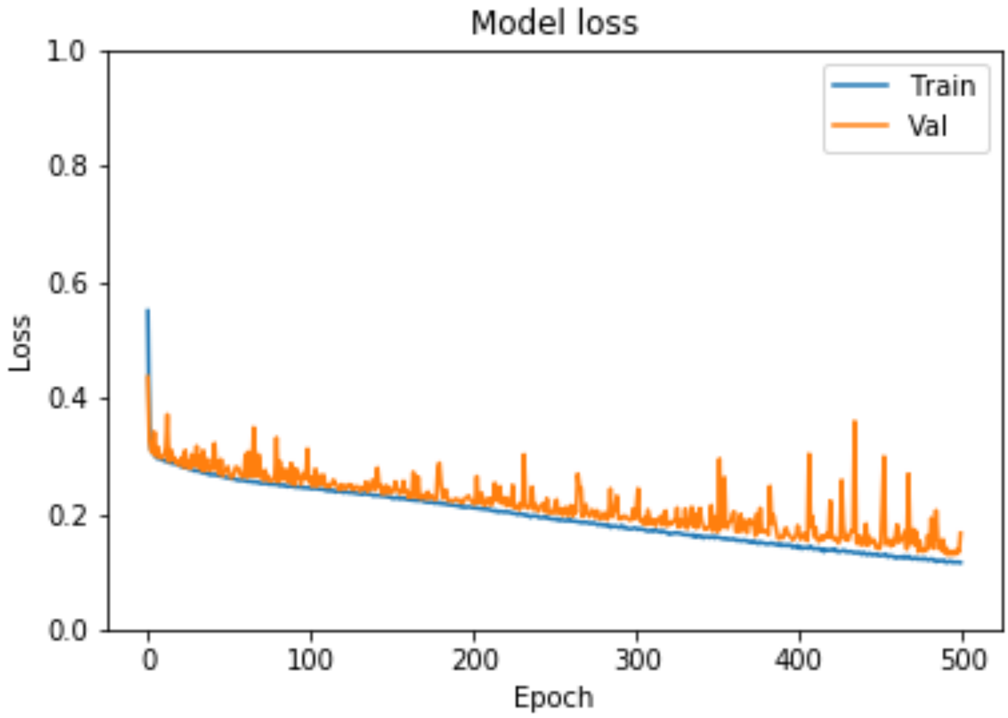

Figure 69: 1D-CNN model loss with HL 2 KS5 dropout .25

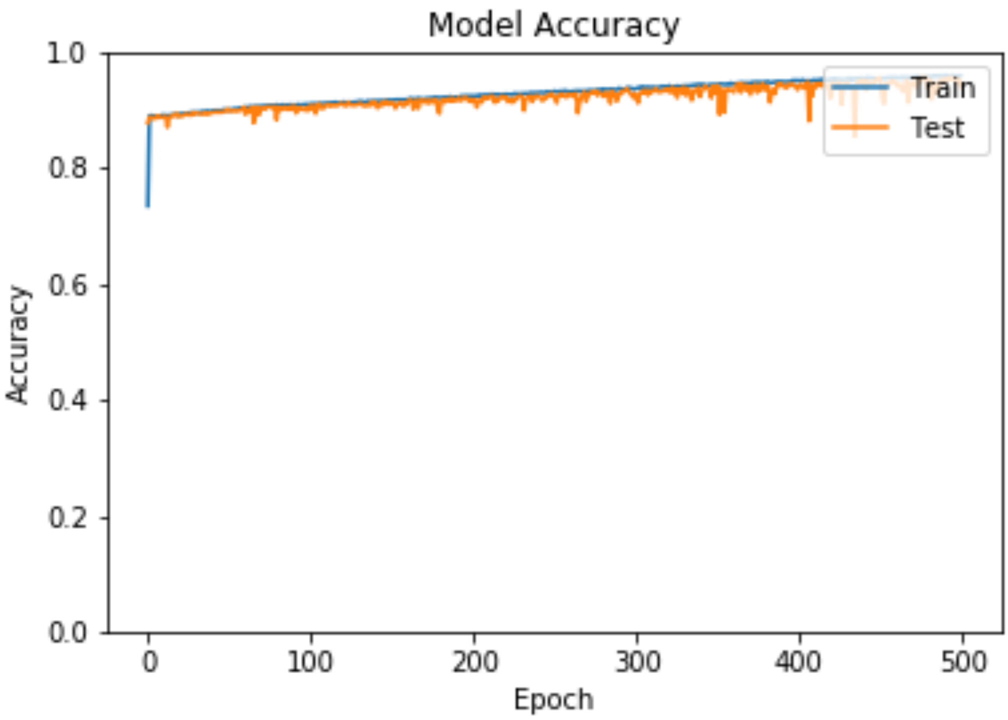

Figure 70: 1D-CNN model accuracy with HL 2 Kernel size 5 dropout .25

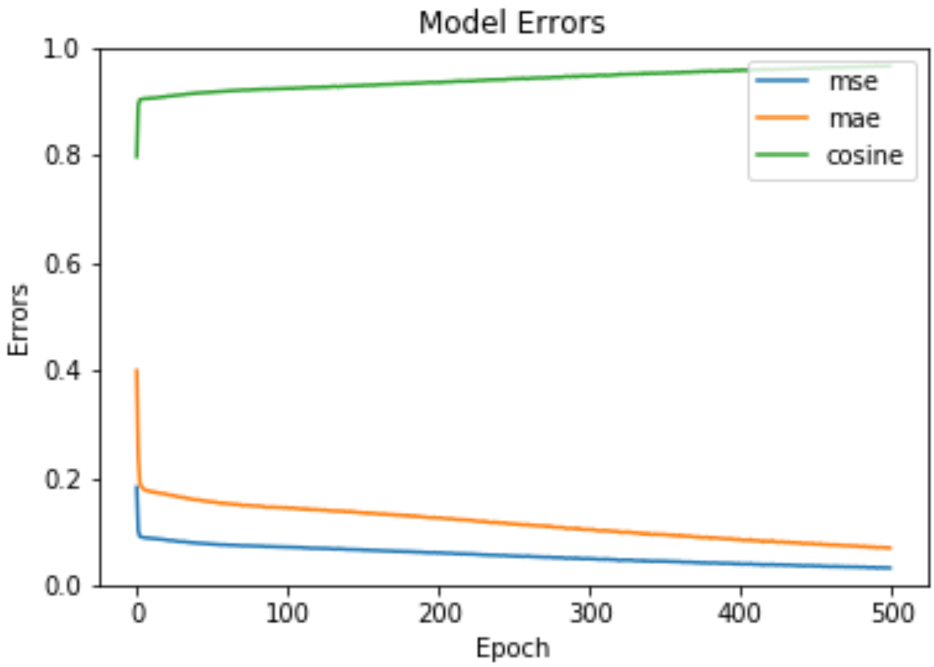

Figure 71:1D-CNN model errors with HL 2 KS5 dropout .25

5.3.5 DL Model with No of Hidden Layers 3 Kernel Size 3 Dropout 0.25

To improve the model's performance, a new model with the number of hidden layers 3, kernel size 3 and dropout .25 was designed. The same data set was used to run the experiment. This model's training accuracy was in the range of 67.39% and 95.78% and the median at 92.87%. The test accuracy was in the range of 32.75% to 94.93%, with a median of 91.68%. Figure 72, Figure 73 and Figure 74 show loss, accuracy and errors of the model. Minimum train accuracy went down, but the upper side went up marginally. A similar behaviour of the model was observed in the test accuracy.

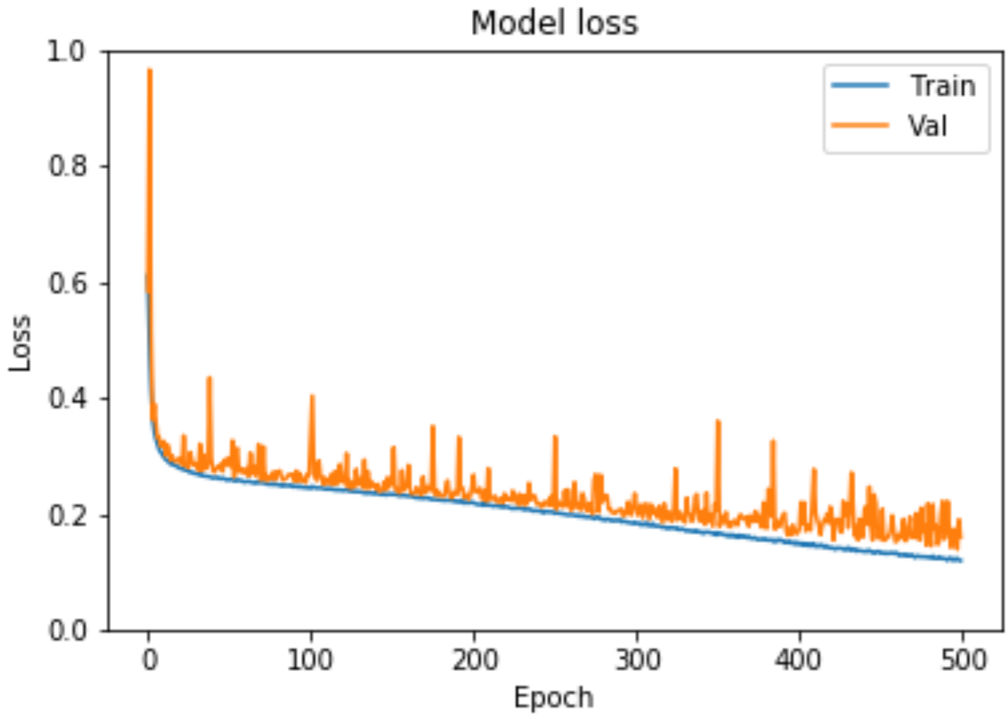

Figure 72: 1D-CNN model loss with HL 3 KS3 dropout .25

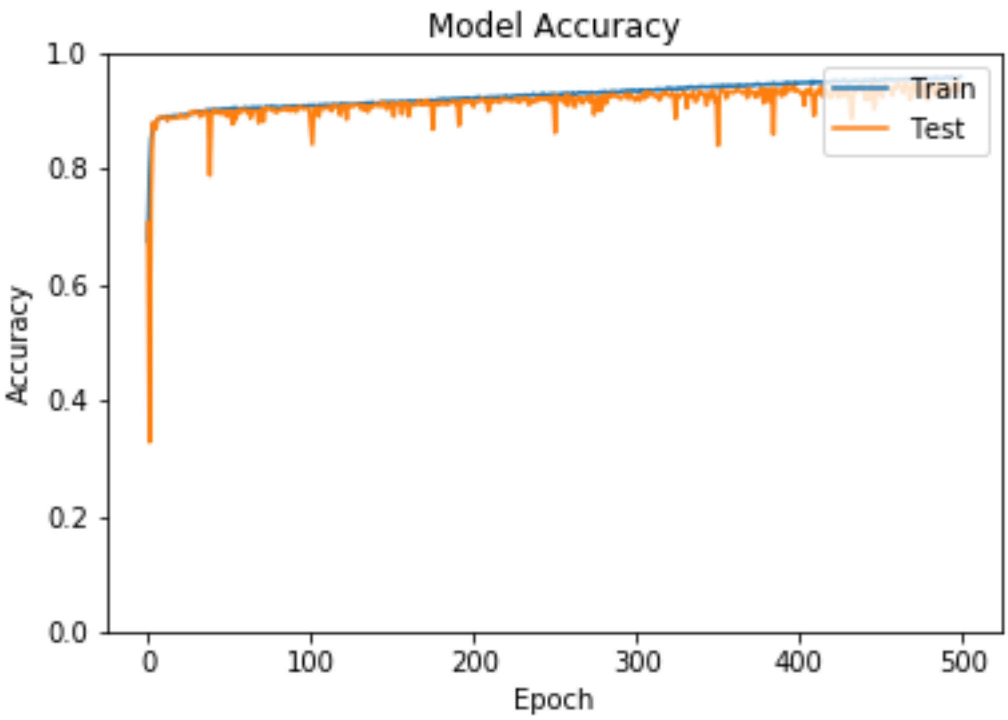

Figure 73: 1D-CNN model accuracy with HL 3 KS3 dropout .25

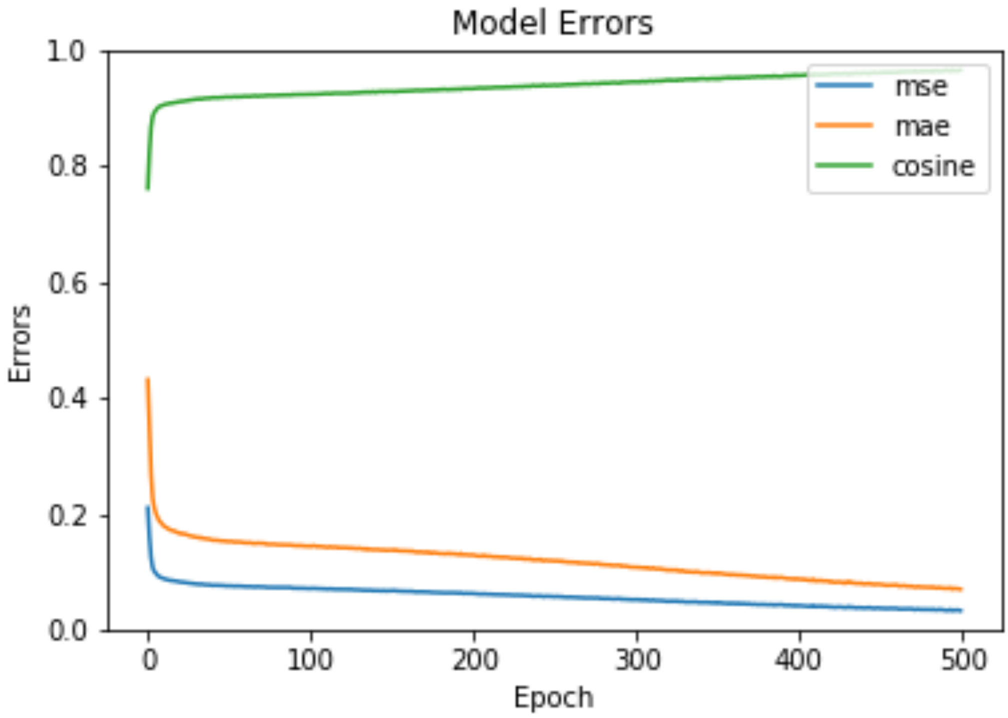

Figure 74:1D-CNN model errors with HL 3 KS3 dropout .25

5.3.6 DL Model with No of Hidden Layers 3 Kernel Size 5 Dropout 0.25

This model has three hidden layers, kernel size 5 and dropout .25. This model's training accuracy was in the range of 69.41% to 96.31% and the median at 93.22%. The test accuracy was in the range of 85.00% to 95.35%, with a median of 91.77%. Figure 75, Figure 76 and Figure 77 show loss, accuracy and errors of the model. Minimum train accuracy went down, but the upper side went up marginally. The same applied to the test accuracy.

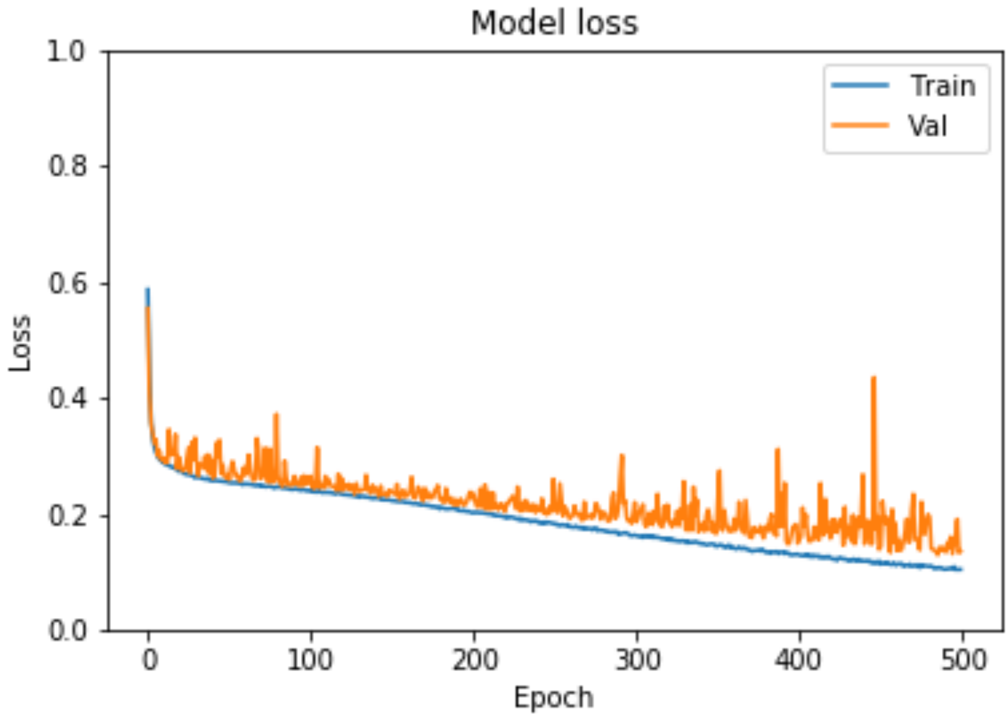

Figure 75: 1D-CNN model loss with HL 3 KS5 dropout .25

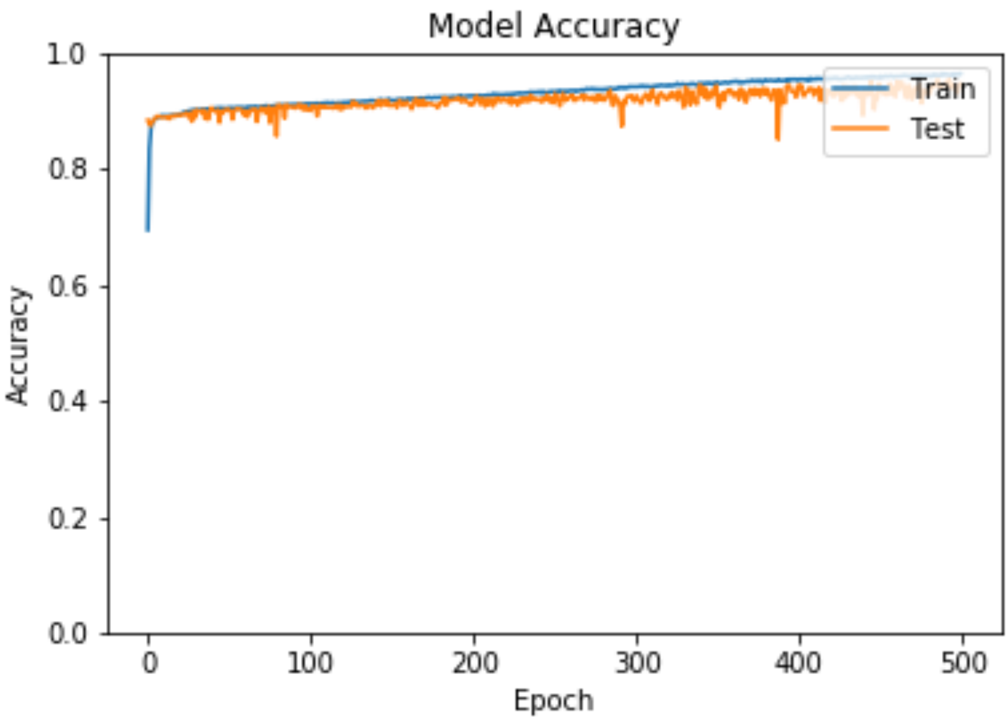

Figure 76: 1D-CNN model accuracy with HL 3 KS5 dropout .25

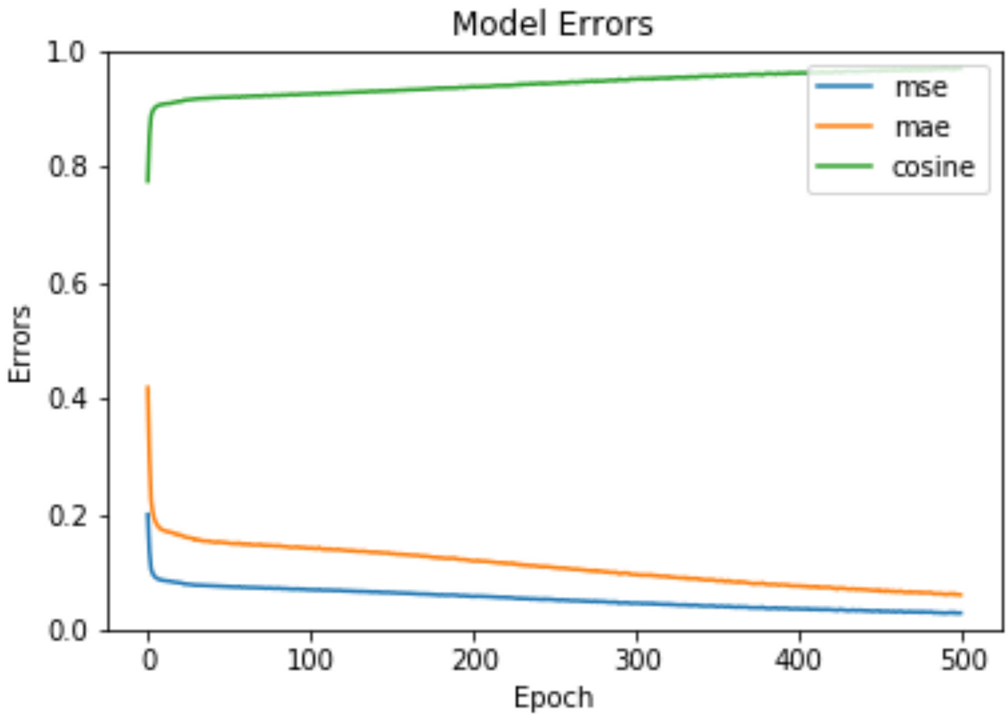

Figure 77:1D-CNN model errors with HL 1 KS5 dropout .25

5.3.7 DL Model with No of Hidden Layers 5 Kernel Size 3 Dropout 0.25

In this model the number of hidden layers was increased to five, kernel size 3 and dropout .25. This model's training accuracy was in the range of 67.18% to 94.51% and the median at 92.15%. The test accuracy was in the range of 49.82% to 92.84%, with a median of 90.43%. Figure 78, Figure 79 and Figure 80 show loss, accuracy and errors of the model. For this model, accuracies were lower than the previous model.

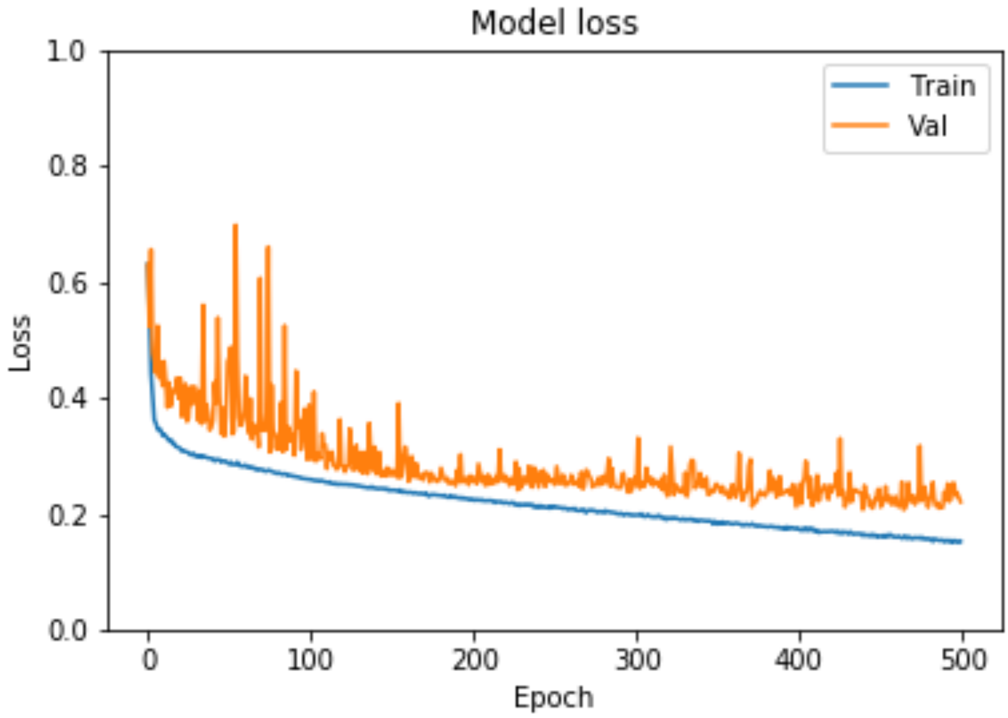

Figure 78: 1D-CNN model loss with HL 5 KS3 dropout .25

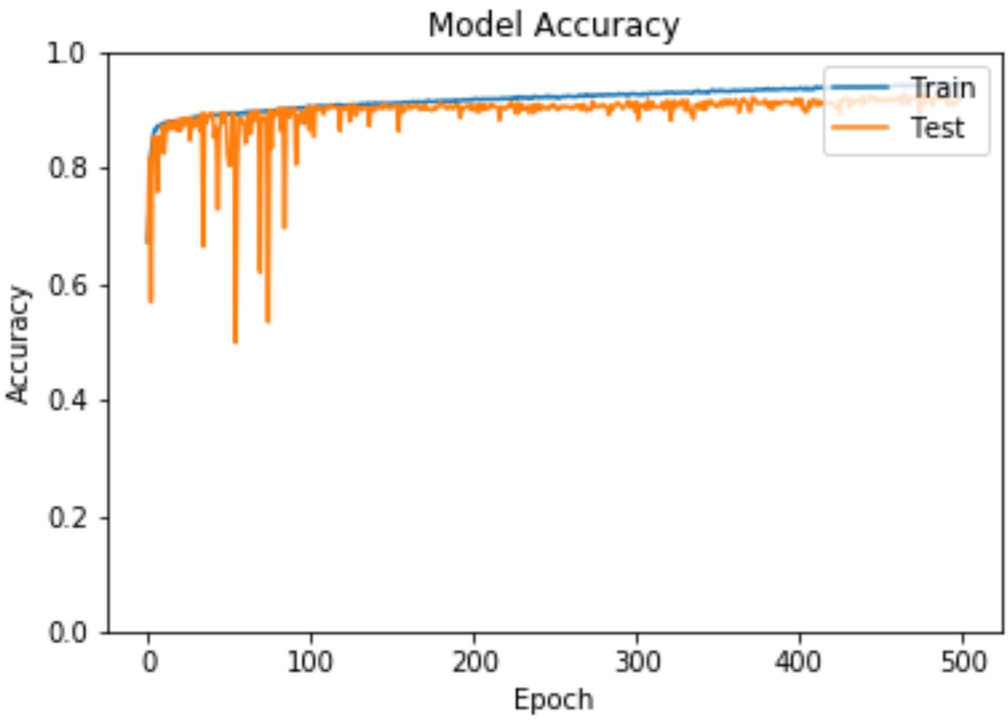

Figure 79: 1D-CNN model accuracy with HL 5 KS3 dropout .25

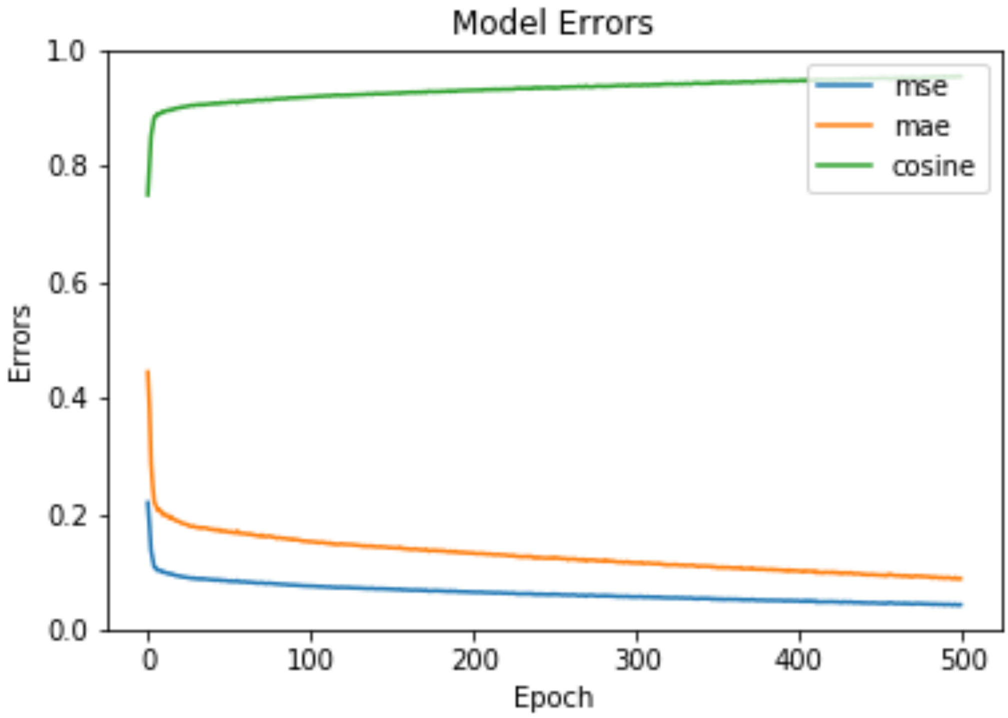

Figure 80:1D-CNN model errors with HL 5 KS3 dropout .25

5.3.8 DL Model with No of Hidden Layers 5 Kernel Size 5 Dropout 0.5

In this model, the number of hidden layers kept the same as in the previous model, but kernel size was changed to 5 and drop out .5. This model's training accuracy was in the range of 66.95% to 92.18% and the median at 90.89%. The test accuracy was in the range of 67.25% to 91.40%, with the median at 89.85%. Figure 81, Figure 82 and Figure 83 show loss, accuracy and errors of the model. For this model, accuracies were lower than the previous model.

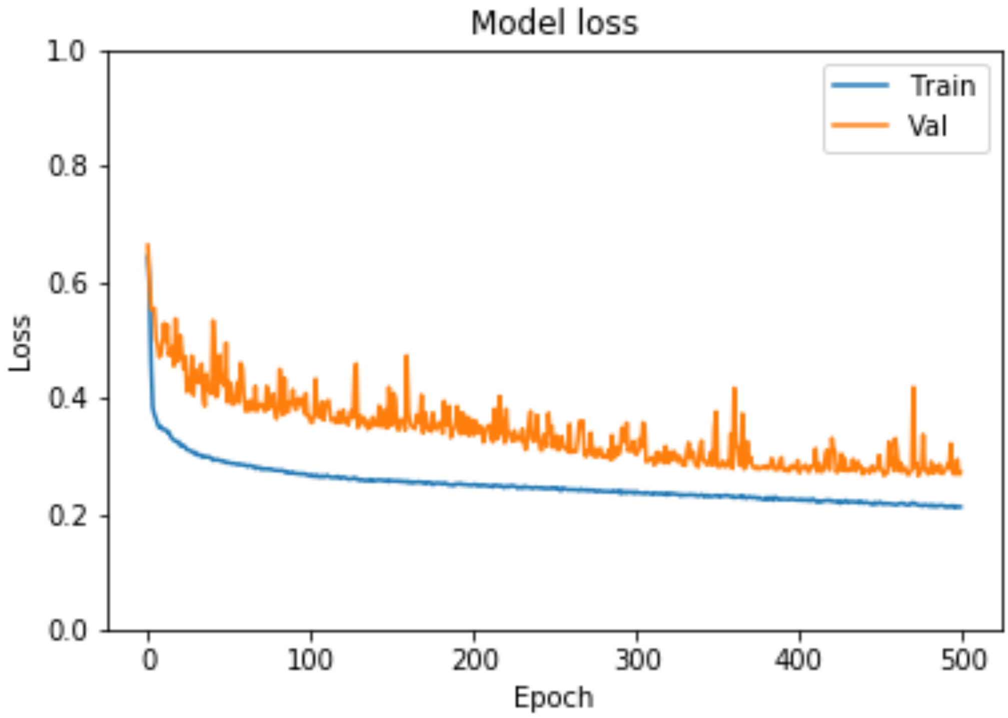

Figure 81 :1D-CNN model loss with HL 5 KS5 dropout .5

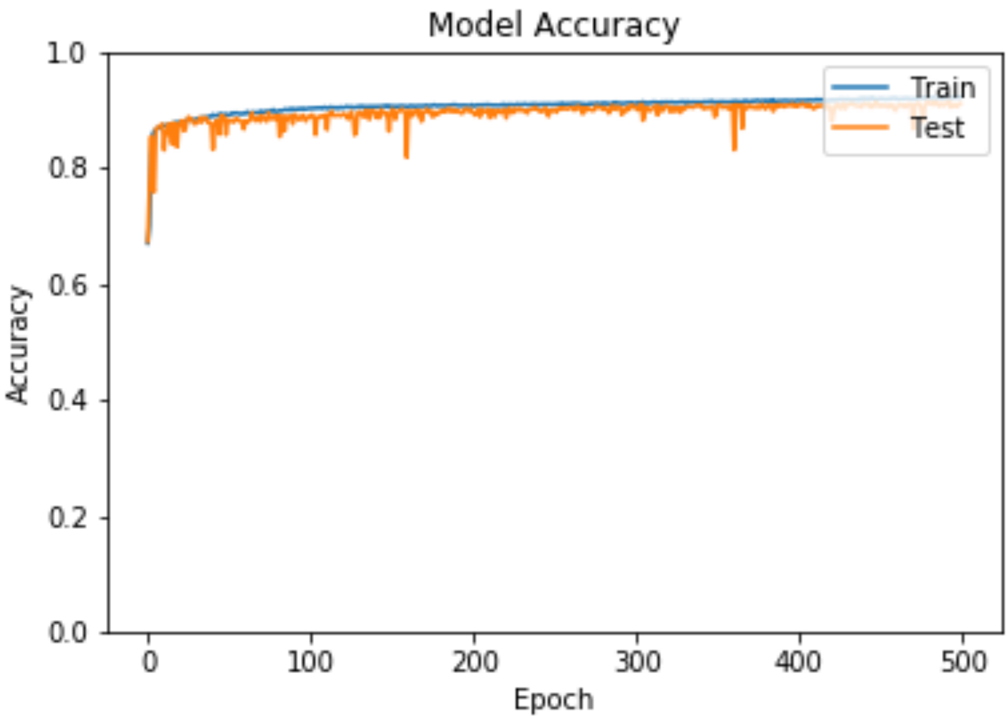

Figure 82: 1D-CNN model accuracy with HL 5 KS5 dropout .25

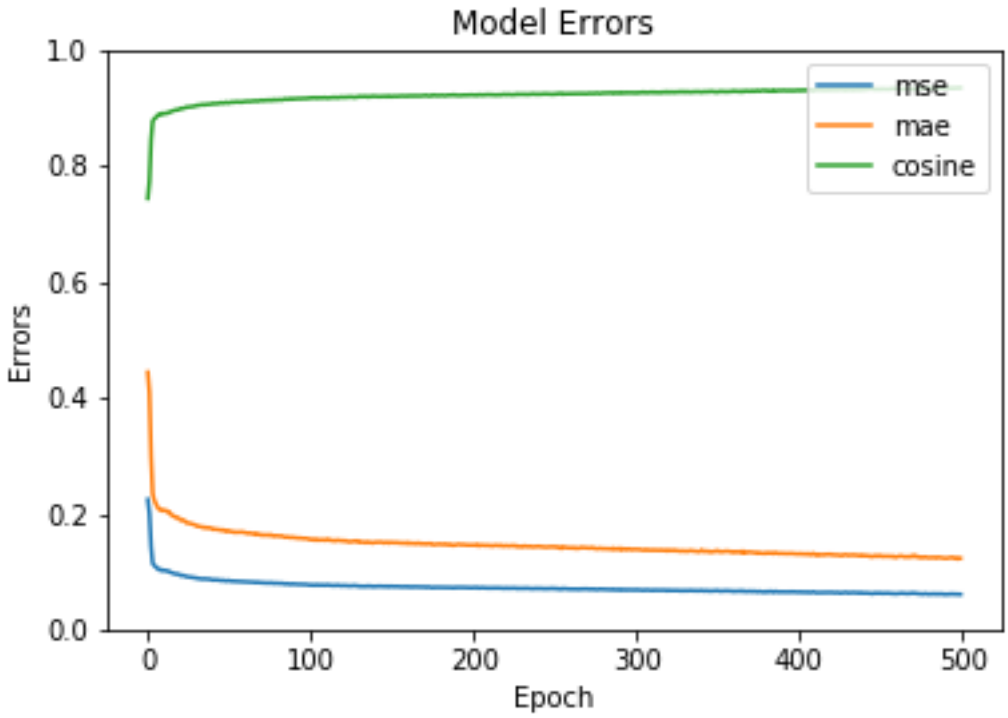

Figure 83:1D-CNN model errors with HL 5 KS5 dropout .25

5.3.9 DL Model with No of Hidden Layers 5 Kernel Size 5 Dropout 0.25

In this model the number of hidden layers and the kernel size were kept the same as in the previous model, but the dropout was 0.25. This model's training accuracy was in the range of 67.20% to 95.77% and the median at 92.82%. The test accuracy was in the range of 63.25% to 93.75%, with the median at 90.76%. Figure 84, Figure 85 and Figure 86 show loss, accuracy and errors of the model. For this model, accuracies were lower compared to the previous model.

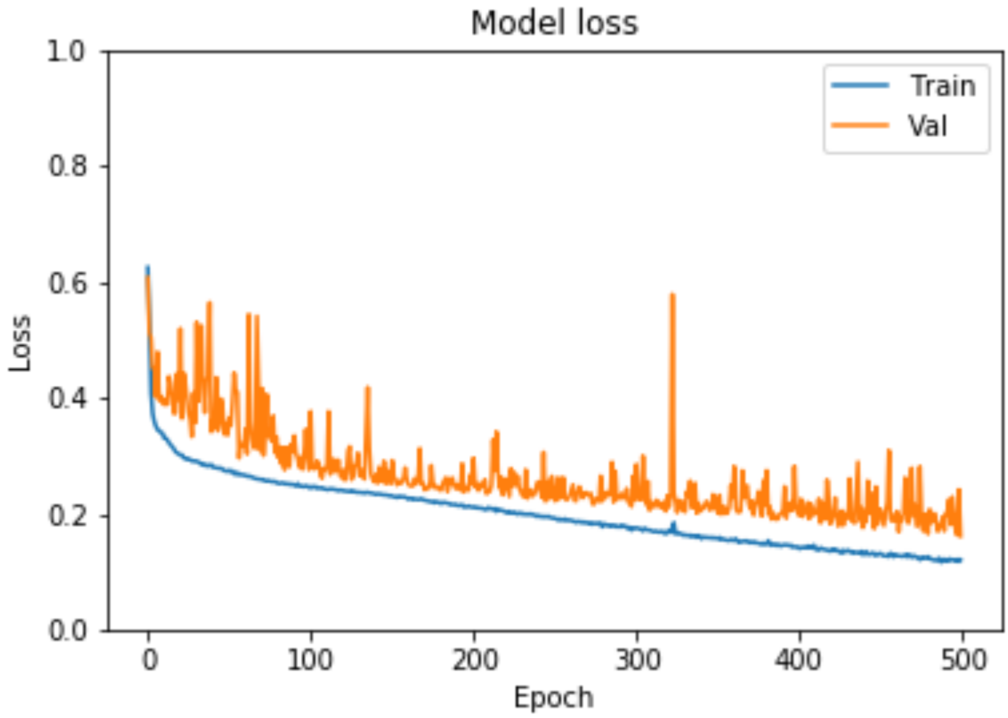

Figure 84: 1D-CNN model loss with HL 5 KS5 dropout .25

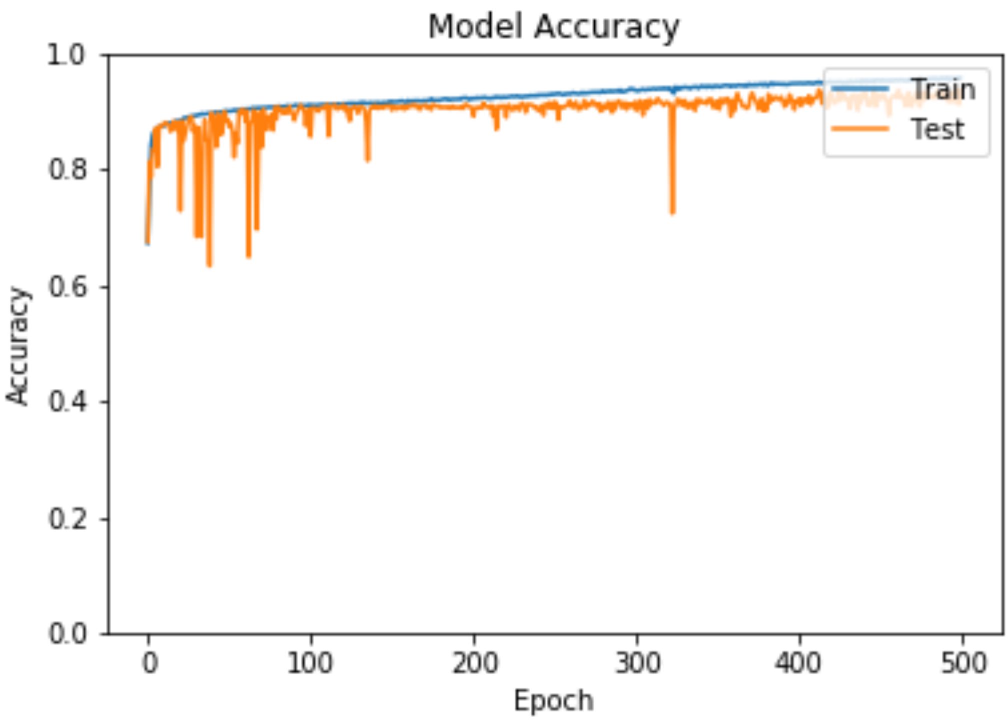

Figure 85: 1D-CNN model accuracy with HL 5 KS5 dropout .25

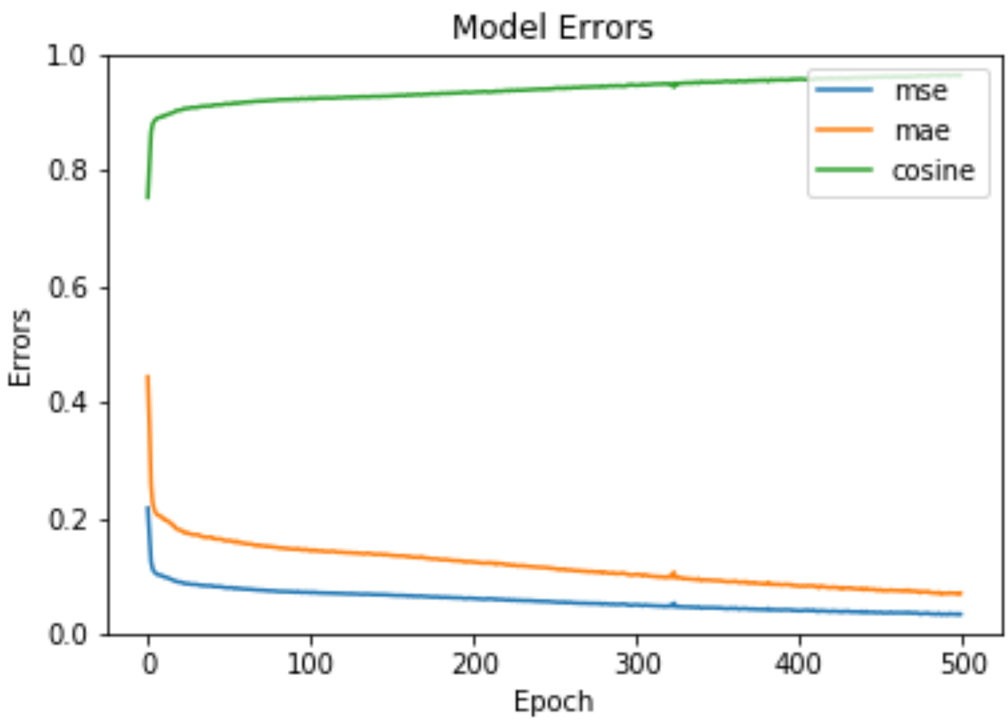

Figure 86:1D-CNN model errors with HL 5 KS5 dropout .25

5.3.10 DL Model with No of Hidden Layers 6 Kernel Size 3 Dropout 0.25

In this model, the number of hidden layers was increased to 6 and kernel size was set at 3 with dropout at 0.25. This model's training accuracy was in the range of 67.17% to 93.58% and the median at 91.78%. The test accuracy was in the range of 41.13% to 92.25%, with a median of 90.38%. Figure 87, Figure 88 and Figure 89 show loss, accuracy and errors of the model. For this model, both train and test accuracies were lower than the previous model. It can be observed that with the increasing number of hidden layers, accuracies of the models are going down.

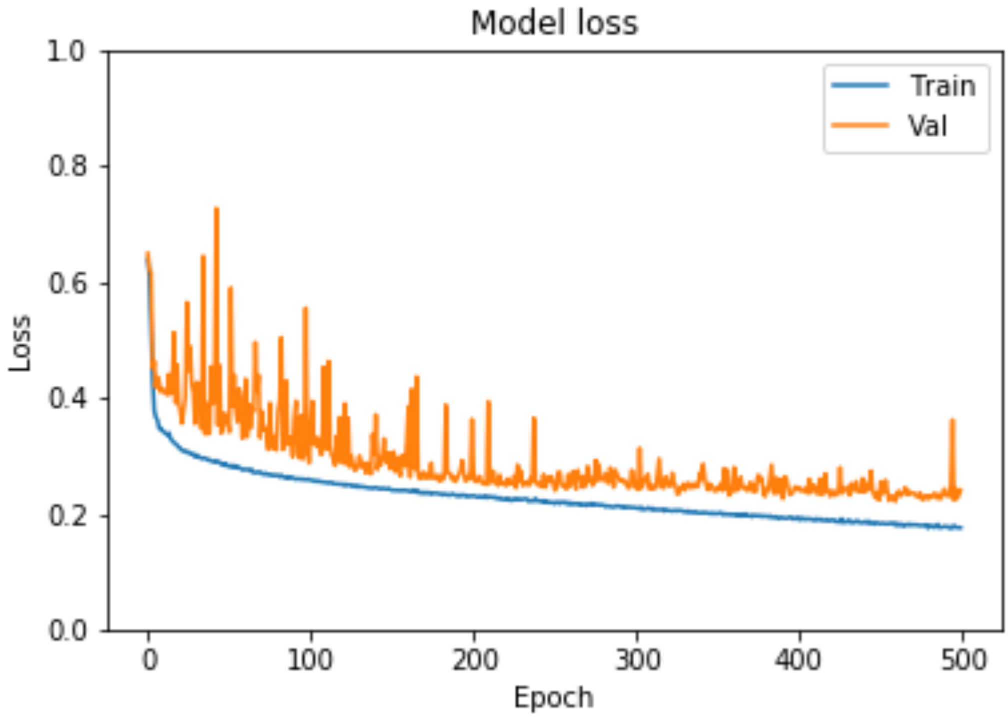

Figure 87: 1D-CNN model loss with HL 6 KS3 dropout .25

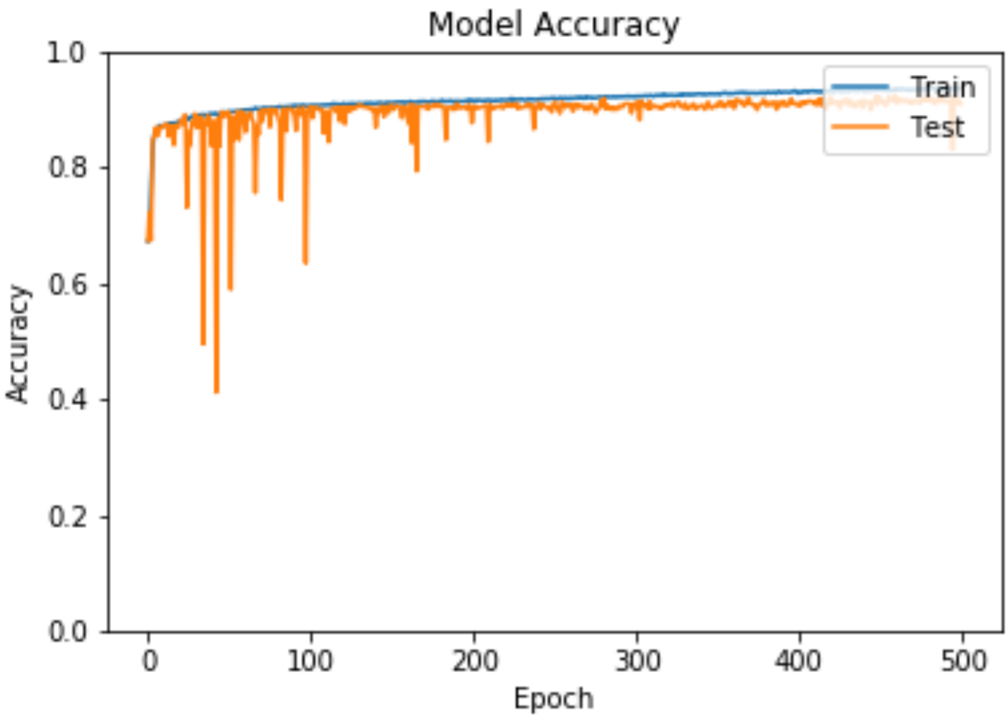

Figure 88: 1D-CNN model accuracy with HL 6 KS3 dropout .25

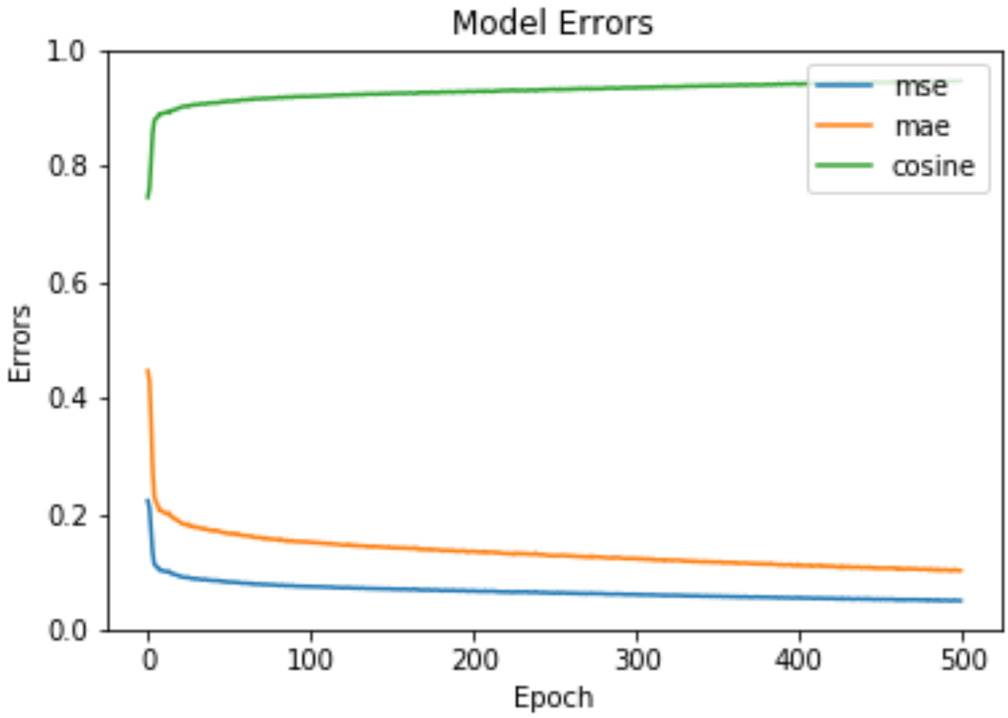

Figure 89:1D-CNN model errors with HL 6 KS3 dropout .25

5.3.11 DL Model with No of Hidden Layers 7 Kernel Size 5 Dropout 0.25

In this model, the number of hidden layers were increased to 7 and kernel size was set at 5 with dropout at 0.25. This model's training accuracy was in the range of 65.93% and 94.88% and the median at 92.35%. The test accuracy was in the range of 55.05% to 93.03%, with a median of 90.45%. Figure 90, Figure 91 and Figure 92 show loss, accuracy and errors of the model. For this model, accuracies were lower than the previous model. It can be observed that with the increasing number of hidden layers, accuracies of the models are going down.

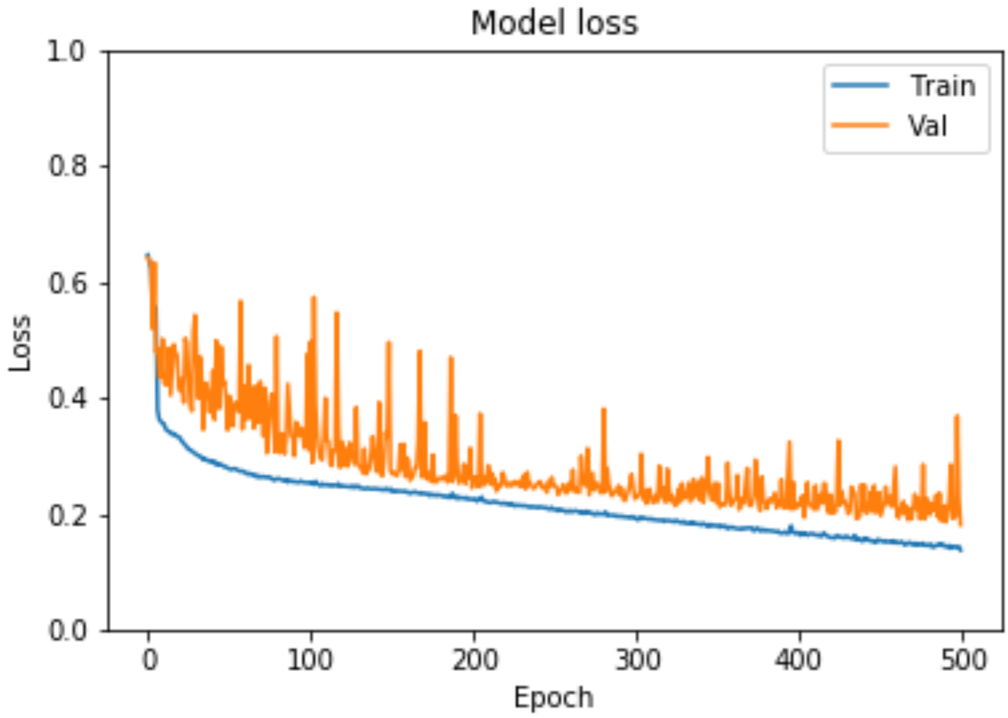

Figure 90: 1D-CNN model loss with HL 7 KS5 dropout .25

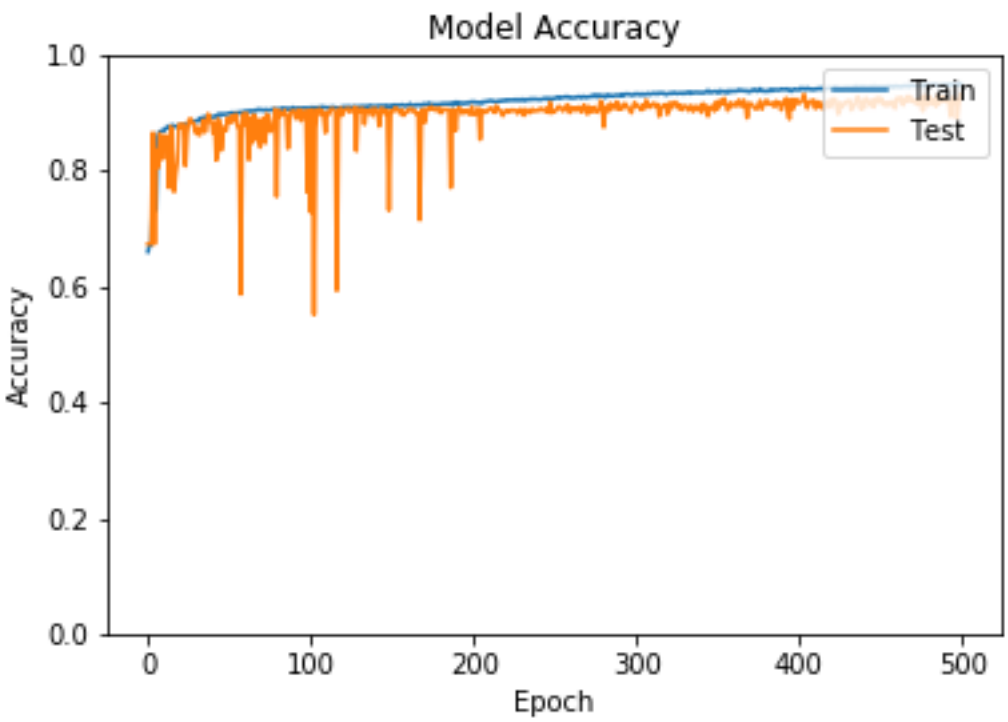

Figure 91: 1D-CNN model accuracy with HL 7 KS5 dropout .25

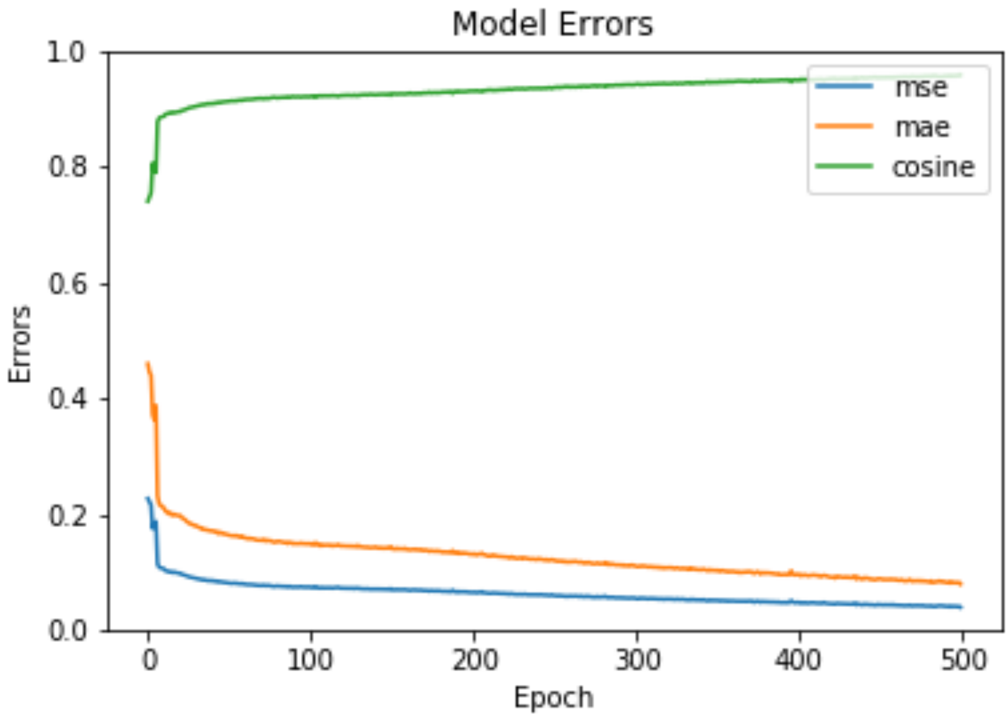

Figure *92*:1D-CNN model errors with HL 7 KS5 dropout .25

5.3.12 Result

Figure 19 shows the steps involved in training the 1D-CNN model, and the algorithm Table 10 gives details of the training process. shows steps to identify a sample using the 1D-CNN model, and the algorithm Table 10 provides details of how it was executed. The data set Table 4 was split into 70% for training, 15% for validation and 15% for testing, then was used to run simulations with various combinations of hyperparameters. Rectified Linear Unit (Relu) activation function was used during the experiments. Each training was conduction with batch size 100 and number of epochs 500. The analysis involved a comparison of test accuracy, precision and recall for various models.

Number of hidden layers	Model Accuracy	Average Precision Rates	Average Recall Rates	F1-Score
7	93.66%	99.00%	73.00%	84.00%
6	93.54%	97.00%	74.00%	84.00%
5	94.40%	99.00%	77.00%	86.00%
3	95.78%	99.00%	80.00%	89.00%
2	95.98%	94.00%	90.00%	92.00%
1	95.10%	85.00%	92.00%	88.00%

Figure 93:Results for 1D-CNN implementation for kernel size=3

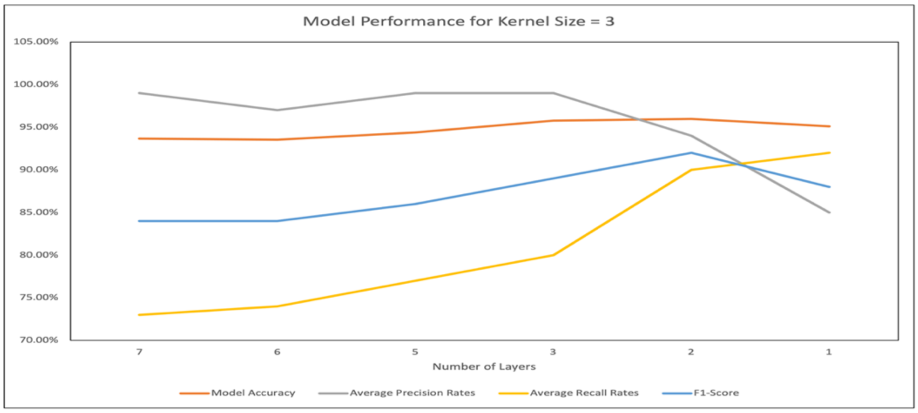

Figure 94: Model Performance for Kernel Size=3

Figure 93 and Figure 95 show the simulation results of the 1D-CNN models with kernel size three and a number of hidden layers ranging from 1 to 7. The model's accuracy was in a small range of around 95%. However, the recall rate and F1 score were better for the model with fewer hidden layers. The recall rate was highest for the model with just one hidden layer, and F1 score was highest for the model, which has two hidden layers. It can be overserved that time to train and classify go down as the number of hidden layers goes down. Looking at all the performance factors discussed above, the model with two hidden layers is the best model with kernel size three.

Number of hidden layers	Model Accuracy	Average Precision Rates	Average Recall Rates	F1-Score
7	94.87%	99.00%	79.00%	88.00%
6	95.43%	97.00%	83.00%	89.00%
5	95.63%	98.00%	81.00%	89.00%
3	96.29%	98.00%	84.00%	91.00%
2	95.74%	89.00%	93.00%	91.00%
1	95.10%	98.00%	81.00%	89.00%

Figure 95: Results for 1D-CNN implementation for kernel size=5

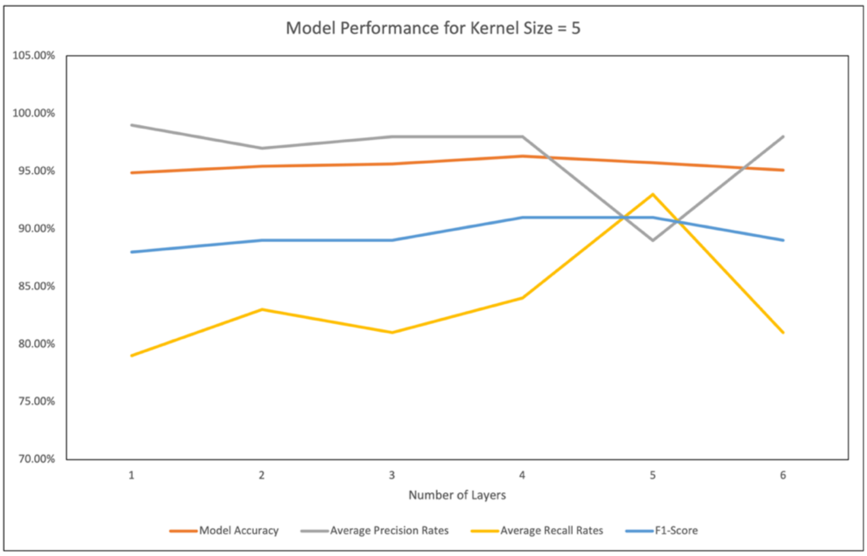

Figure 96: Model Performance for Kernel Size 5

Figure 95 and Figure 96 show the simulations results of the 1D-CNN models with Kernel size five and the number of hidden layers ranging from 1 to 7. As discussed above for the models with kernel size three, the accuracy of the model with kernel size five was also in a small range of 94.87% to 96.29%. The recall rate and F1 score were better for the model with the fewer number of hidden layers. The recall rate was highest for the model with two hidden layers, and then it went down for the model with one hidden layer. The F1 score is in the range of 88%-91% and is highest for the model, which has two and three hidden layers. Looking at all the performance factors discussed above, the model with two hidden layers might be the best model with kernel size three. So, looking at various the results it can be concluded that, one dimension Convolution Neural Networks model with two layers and kernel size of 5 be sufficient for the dataset described in Table 4.

Using a deep neural network-based approach enabled our study to achieve better accuracy compared to the work of (Astarita & Vittoria, 2012). In (Astarita & Vittoria, 2012),the researchers used sensory data and applied filters to isolate potholes and achieved a pothole detection rate of 65%. Moreover, the recall rate, precision and F1-Score in this study were better compared to the study (Chao , et al., 2020) that utilized simpler machine learning approaches such as Linear Regression, Support Vector matrix and Random Forests to perform the classification tasks.

5.4 Fusion DL Model to Detect Road Pothole Using Sensory and Imagery Data

There are three main layers in Convolution Neural Networks architecture: Convolution Layer, Pooling Layer and Fully connected Layer. The Convolution Neural Networks reduce the model's learning complexity (Ahmed & Tao, 2017) by sharing the weights during training. The model's capacity and complexity can be changed by changing the number of Convolution Neural Networks layers and their organization. This study used a fusion of two-dimension Convolution Neural Networks (CNN) and one-dimensional Convolution Neural Networks, as shown in Figure 97. One branch, 1D-CNN, takes accelerometer data as input and another branch, 2D-CNN, takes images as input. Later down the layer hierarchy of Deep Neural Networks (DNN) outputs of 1D-CNN and 2D-CNN are concatenated, and a dense layer is used to produce a final result. The 2D-CNN models with hidden layers of non-linear transformation ranging from three to five were used to obtain the classification results. 1D-CNN models with multiple non-linear transformation layers ranging from two to seven were used to get the classification results. It was observed that the 2D-CNN model with five hidden layers and 1D-CNN networks with two hidden layers had produced satisfactory results. This study has used a fusion model of 2D-CNN with five hidden layers and 1D-CNN with two hidden layers. The activation function on hidden layers was ReLu (Rectified Linear Unit), and on the output layer, activation function SoftMax was used. The loss function was categorical cross-entropy (Eq:1) and Adam optimizer used.

$$L(X,Y;\theta) = -\frac{1}{N}\sum_{1}^{N}(-y_p \,.\log(\hat{y}_{i-p}) - y_n.\log(\hat{y}_{i-n})) \text{ --Eq. 5}$$

$$\theta = \theta - \eta.\nabla_{\theta}L(x^i, y^i, \theta) \qquad \text{---Eq. 6}$$

Where θ is a weight parameter, the training aims to minimize loss (Eq.1) and get optimal weight parameter θ. For the samplex_i , the predicted negative probability is denoted by$\hat{y}_{i-n}$, and positive probability by$\hat{y}_{i-p}$

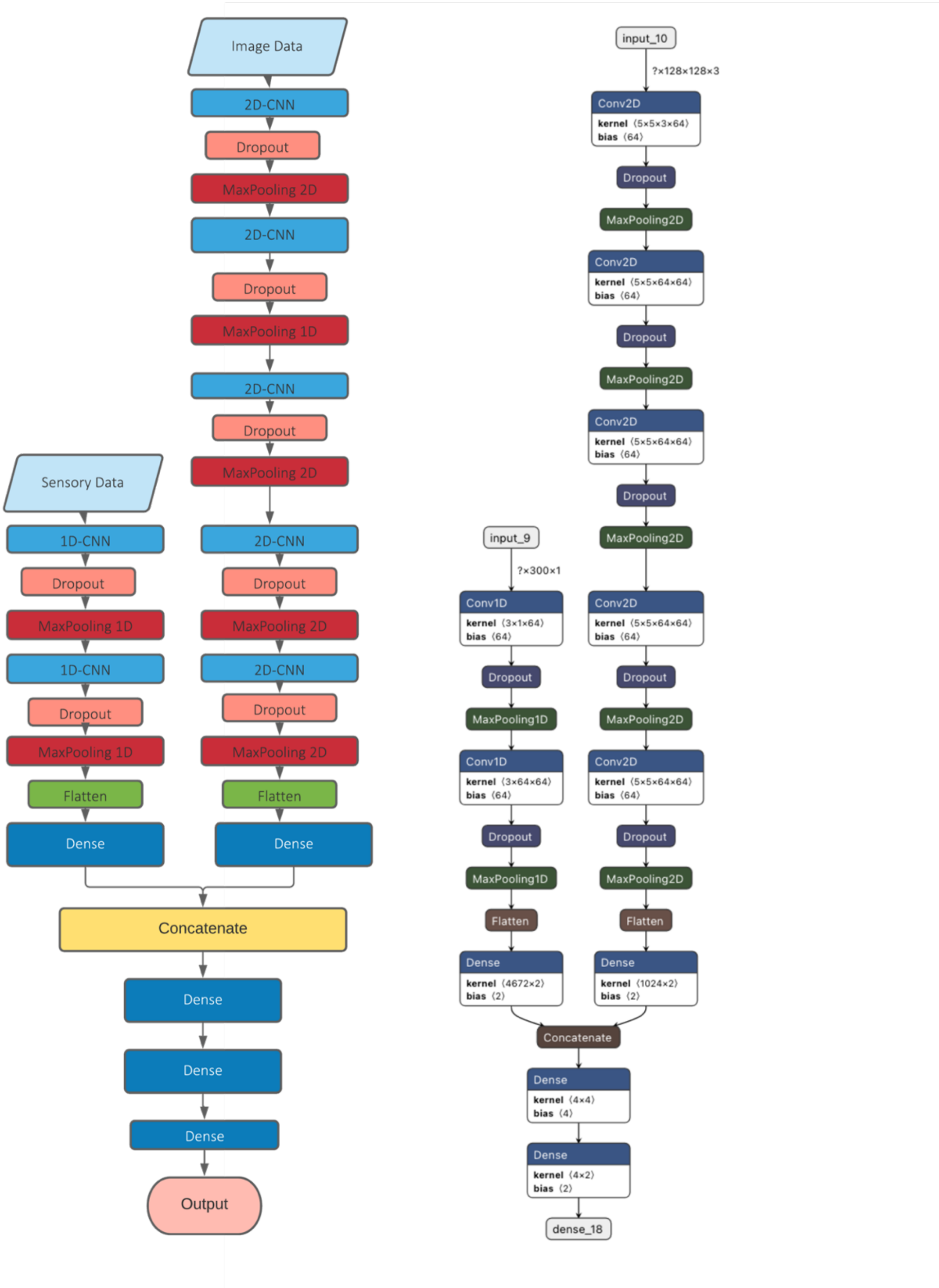

Figure 97: Fusion convolutional neural network

Table 11 :Road pothole identification process using fusion convolution neural network

Algorithm : Road Pothole Identification process using fusion Convolution Neural Network
Input : Sensory data and Imagery Data Output: Pothole or No pothole; GPS location • Capture sensor data using an accelerometer sensor app and image using the camera of the smartphone • Upload data on a cloud server • Download data from server to a computer • Pre-process accelerometer data to make that one-sec data sample has 100 timestamps(on all three axis X, Y, Z); if not, do resampling • Crop area of interest in the images • Feed data from the previous step to the classifier model • If$y_p > y_n$ then • The fed sample represents a pothole • Else • No pothole (normal road surface) • End if • Tag the sample with pothole and GPS details. • Upload the classifier on the cloud server • Send notification to other vehicles(who have opted for) which are in proximity.

The dataset Table 4 was split into 70% for training, 15% for validation and 15% to test the fusion model. The input size of the images was 128x128. Each experiment was run for 50 epochs on computer Table 9.

The design of fusion model was decided based on the performance of one-dimension Convolution Neural Networks model for sensory data and two-dimension Convolution Neural Networks model for imagery data. It was observed that the best performing one-dimension Convolution Neural Networks model was with 2 hidden layers, kernel size 5 and dropout .25. The best performing two-dimension Convolution Neural Networks model for imagery data was with 5 hidden layers. Few fusion models with varying hidden layers were tried as below.

1DCNN = one Dimension Convolution Neural Networks

2DCNN = Two Dimension Convolution Neural Networks

5.4.1 Fusion Model 1DCNN-2L-2DCNN-2L

The first fusion model has two hidden layers in the 1D-CNN branch for sensory data input and two hidden layers in the 2D-CNN branch for imagery data input. The training accuracy was in the range of 68.55% to 81.91% and the median at 81.65%. The test accuracy was in the range of 75.89% to 82.84%, with a median of 80.84%. Figure 98 shows training and test accuracies for this model.

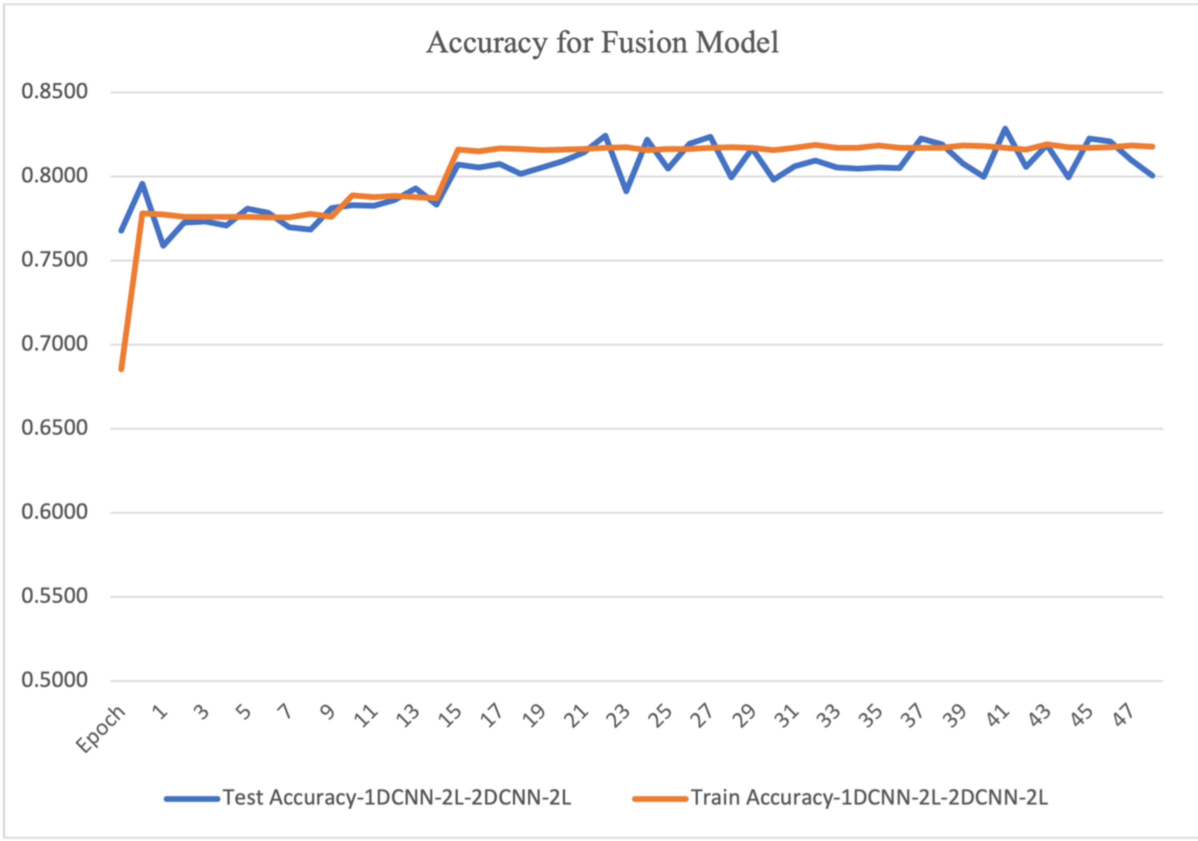

Figure 98:Accuracy fusion model 1D-CNN HL 5 -2D-CNN HL 2

5.4.2 Fusion Model 1DCNN-2L-2DCNN-5L

After the first fusion model, as discussed above, a new fusion model with 2 hidden layers in the 1D-CNN branch for sensory data and 5 hidden layers in the 2D-CNN branch for imagery data was used to train on the same data set as in the previous experiment. Figure 99 shows training and test accuracies of this mode. The training accuracy was in the range of 74.28%

and 96.97% and the median at 96.50%. The test accuracy was in the range of 80.62% and 96.91%, with a median of 94.49%. It can be observed that the model has stable and smooth accuracies.

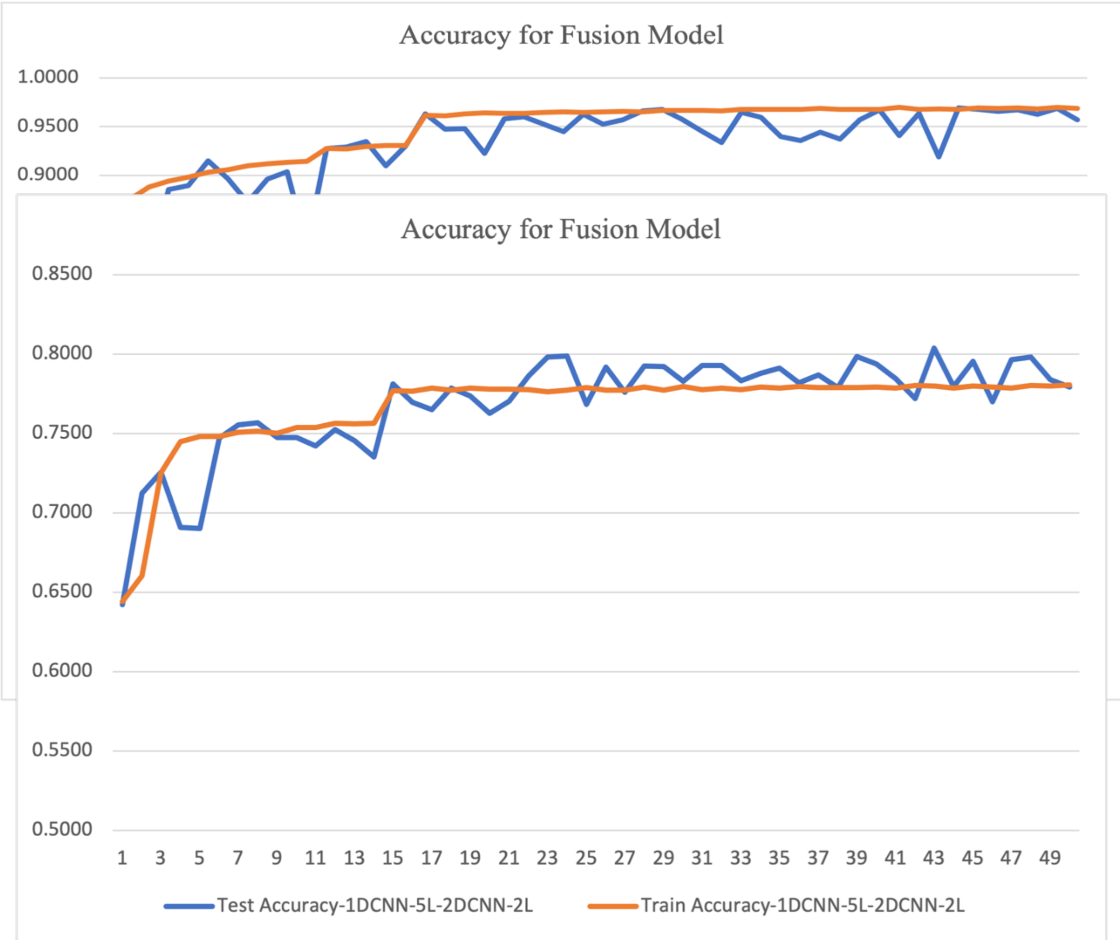

Figure 99:Accuracy fusion model -1D-CNN-HL 2-2D-CNN-HL 5

5.4.3 Fusion Model -1DCNN-5L-2DCNN-2L

The last fusion model used in this study has 5 hidden layers in the 1D-CNN branch for sensory data, and 2 hidden layers in the 2D-CNN branch for imagery data was used to train on the same data set as in the previous experiment. The training accuracy was in the range of 64.41% and 78.07% and median 77.78%. The test accuracy was in the range of 64.21% and 80.37%, with a median of 77.90%. The performance of this model was not as good as the immediately previous model.

Figure 100:Accuracy Fusion Model -1D-CNN HL 5 -2D-CNN HL2

5.4.4 Results

After analysis of 1D-CNN models and 2D-CNN models, it was decided to design a fusion model that can take advantage of both imagery data and sensory data. Few fusion models were designed and trained, as discussed in section 5.4. One of the many fusion models used in the experiments is shown in Figure 97 . To simulate proposed fusion Convolution Neural Networks, three combinations of hidden layers in 1D-CNN and 2D-CNN were used. The first combination is with two hidden layers in both 1D-CNN and 2D-CNN (1D-2L-2D-2L); the second combination is five layers in the 1D-CNN branch, two layers in 2D-CNN branch, and the last combination has two layers in 1D-CNN and five layers in 2D-CNN. Figure 101 shows training and test accuracies of the fusion models, which were tried in 5.4.1,5.4.2 and 5.4.3.

It can be noted that the fusion model with two hidden layers in the 1D-CNN branch and five hidden layers in the 2D-CNN input branch has the highest validation and training accuracy, 95.71% and 96.87% respectively. The accuracy of the model will increase with the size of the data set and when potholes' type size, shape and depth have more variation.

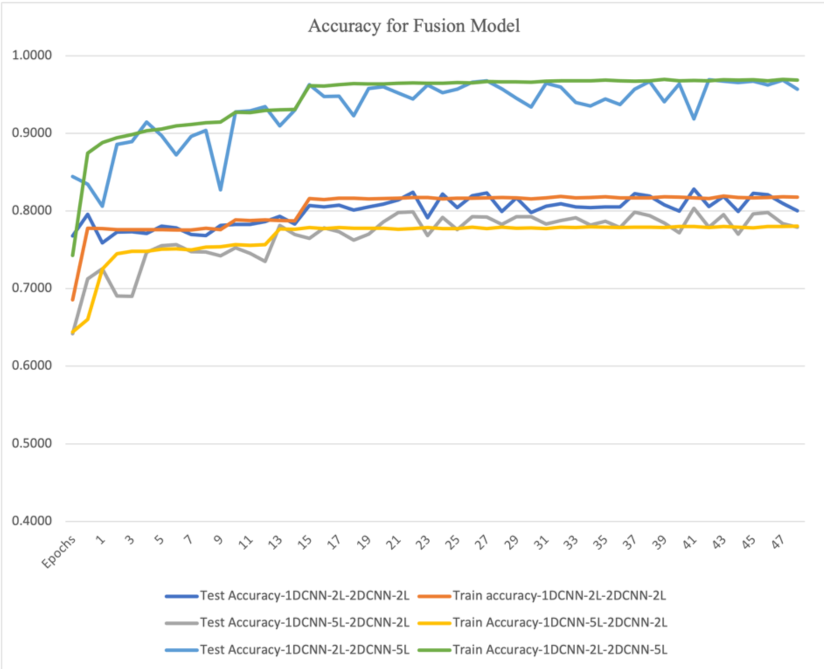

Figure 101: Accuracies of fusion models

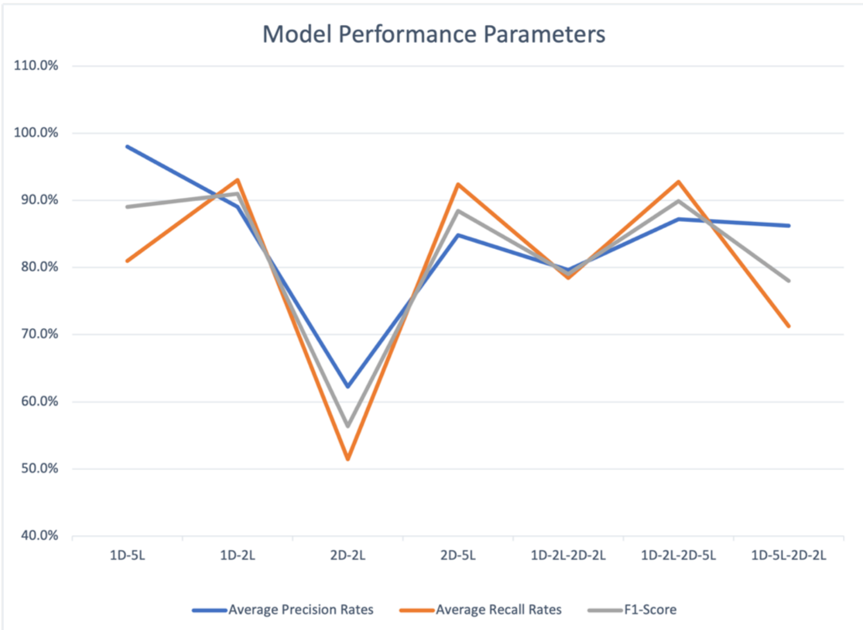

Figure 102: Fusion models performance accuracies

5.5 Chapter Summary

This chapter has discussed the development of a 2D-CNN model and its application for image processing. The state-of-the-art 2D-CNN model was developed for pothole detection. The results show that the model (with 5 layers and 2 classes) has achieved Average Recall Rate (ARR) -92.40% and Average Precision Rate (APR) 84.80%.

Secondly, this chapter has discussed the development of a 1D-CNN model to process sensory data for pothole detection. The model was developed based on a state-of-the-art deep learning-based approach. Various combinations of layers and hyperparameters were used to design 1D-CNN. The model has achieved Average Recall Rate (ARR)-89.00% and Average Precision Rate (APR)-93.00%

Finally, this chapter has discussed the development of a fusion model to process imagery and sensory data for pothole detection. The fusion model is developed based on two deep learning models consisting of 2D-CNN and 1D-CNN. The proposed fusion model takes advantage of the strength of both models and data types (imagery and sensory) and produces better classification results. 2D-CNN was used to process the imagery input while 1D-CNN was used to process sensory input as described above. Both inputs were combined and cross validated to filter out any noise in the data as described above. Confusion matrix was applied to assess the performance of the model. The standard classification performance metrics such as accuracy, precision, recall and loss are derived from the confusion matrix data. The model has achieved Average Recall Rate (ARR) 92.7% and Average Precision -Rate (APR) 87.2%. that constitute better Average Recall when compared to the existing state of the art approaches relating to pothole detection.

The next chapter will conclude this thesis by providing a summary, main contribution and future work.

CHAPTER 6

CONLCUSIONS AND FUTURE WORKS

6.1 Introduction

This thesis has discussed the importance of the real-world issue of pothole detection. Importantly, this thesis has addressed the issue of pothole detection by running a series of experiments using Deep Learning approaches. The Deep Learning approaches were applied to process sensory and imagery data for pothole detection. Most importantly, this long-standing challenge is addressed, through the development of a fusion model based on Convolution Neural Networks.

The rest of the chapter is organised as follows. Section 6.2 provides the thesis summary by highlighting the significance of the research undertaken in this thesis. Section 6.3 presents the contribution of this thesis. Section 6.4 discusses the main limitations of the proposed work. Section 6.5 outlines the future work.

6.2 Thesis Summary

The research documented in this thesis started with the identification of issue relating to pothole detection which led to the development of road pothole detection technique using Machine Learning methods. The modern Machine Learning methods can support automated monitoring of roads and highways to lessen the burden on the corresponding authorities and expedite the overall process. In CHAPTER 2 a comprehensive literature review was conducted to understand the application of Machine Learning for image classification and identification. The literature review also covered works that used sensory data and Machine Learning to monitor road surface and detect road anomalies. After completing the literature review, it was noted that several research efforts have focused on image classification using Machine Learning, and the use of sensory data for detecting road surface anomalies. The perceived limitations of the existing approaches were also discussed in the chapter which revealed the issues concerning computer vision models and sensory data models to detect road potholes. More specifically, the computer vision method, which uses imagery data to detect road anomalies, fails to differentiate between potholes and puddles. The vibration method, which uses sensory data, fails to distinguish between a pothole and other anomalies, such as hinges. In this thesis, these issues are addressed by developing a novel fusion method which can combine the advantages of computer vision methods and sensory data methods. The existing research has not combined both methods in the same manner to apply to two types of data for pothole detection.

This thesis has developed a first comprehensive approach based on Deep Learning to address the issue of pothole detection while overcoming the limitations of the existing methods. The new fusion Deep Learning model comprises of a 2D-CNN model and an 1D-CNN model. The model is able to produce accurate classification results based on sensory and imagery data. Towards the development of a fusion model, 2D-CNN model and 1D-CNN model were designed and trained based on the available data set. Different configurations of 1D-CNN and 2D-CNN models were assessed to test the results. In order to systematically apply the model, a research methodology was developed to design, train and test the Machine Learning model developed in this thesis. The methodology is discussed in detail in CHAPTER 3.

The next step in this research process was to collect imagery and sensory data for the project. The data collection and data pre-processing were critical parts for this study. CHAPTER 4 discusses the various stages of data collection and data pre-processing. In order to collect data, an iOS smartphone (iPhone) was used. Sensory data (accelerometer data) was collected with the use of a generic application installed on the same iOS smartphone. The smartphone was securely placed on the windshield of a vehicle. Following that the data was was pre-processed and annotated in order to train the Deep Learning methods. The imagery data and sensory data were investigated manually; any data sample which was found to be unfit for the experiment was removed. The labelling of images as "pothole" or "no pothole" was done manually. The sensory data was labelled using a python script that matched the date and time of the labelled images to the sensory data. Once imagery and sensory data were labelled into classes, the data set was split into 70% for training, 15% for validation and 15% for testing.

CHAPTER 5 discusses the three experiments carried out to develop, refine and evaluate three Deep Learning Models. In the first experiment, the 2D-CNN Deep Learning model was designed and trained with imagery data. Experiment -1 which covers road pothole detection using imagery data has been discussed in detail in section 5.2. Several 2D-CNN models with varying numbers of hidden layers and different sizes of images were designed and trained as mentioned in sections 5.2.1 to 5.2.5. The results from these simulations were critically analysed and observed that the 2D-CNN model with five hidden layers produced the best result as compared to the other models. The 2D-CNN model has produced 84.80% precision and 92.40% recall, and it was deemed to be effective for pothole detection.

The next step in the study was to design, train and test different one-dimensional Convolution Neural Networks (1D-CNN) models using sensory data and with different hyperparameters (5.3). The designed 1D-CNN models were trained and tested. The results for these models have been discussed in section 5.3.12.The experiments demonstrated that 1D-CNN could effectively detect potholes based on accelerometer data collected from standard smartphones. The results showed that a 2-hidden layer 1-D CNN with kernel size 3 could achieve excellent results. This configuration was able to obtain a precision of 98%, recall of 84%, and F1-score of 91%. These scores correspond to state-of-the-art performance for pothole detection that utilizes sensory data. While its accuracy rate is considerably high, it does not require any special equipment to collect and process the data. Moreover, the raw data that is collected by the smartphone sensor can be used to train the 1D-CNN model without the need for excessive data pre-processing. This fact reduces the computational complexity and the time needed for delivering the classification results in real time.

After completing the experiment with the 1D-CNN model with sensory data and the 2D-CNN model with imagery data, it was noted that there is a possibility to develop a fusion Deep Learning model that uses both imagery and sensory data to address the weaknesses of the previous methods and provide a better and more robust road pothole detection method. So, a fusion model (3.5) was developed. Different fusion models were trained and tested using varying hyperparameters (described in section 5.3). After analysing the results of the fusion models, it was observed that the fusion model with two layers in the 1D-CNN branch, and five layers in the 2D-CNN branch, was the best. The fusion model demonstrated excellent performance with 87.2% precision, 92.7% recall and 89.9% F1-Score.

6.3 Contribution

The thesis has made a number of contributions to knowledge which are described below.

The most notable contribution is the development of a fusion model to process imagery and sensory data for pothole detection. The fusion model is developed based on two Deep Learning models consisting of 2D-CNN and 1D-CNN. The proposed fusion model takes advantage of the strength of both models and data types (imagery and sensory) and produce accurate classification results with 87.2% precision, 92.7% recall. 2D CNN was used to process the imagery input while 1D-CNN was used to process sensory input. Both inputs were combined and cross validated to filter out any noise in the data as described above. The model has achieved better accuracy as compared to the existing state of the art approaches relating to pothole detection.

The second contribution is the development of 2D-CNN model and its application for image processing. The state of the art 2D-CNN model was developed to assess its effectiveness for pothole detection. The perceived limitation and strengths of the 2D-CNN model were carefully investigated and reported in chapter. The model has achieved 84.80% average precision rate and 92.40% average recall rate.

The third contribution of this thesis is to develop 1D-CNN model to process sensory data for pothole detection. The model was developed based on state-of-the-art Deep Learning based approach which has achieved average precision of rate 98%, and average recall rate of 84%. Further strengths and limitations of this model were carefully assessed so that a new model can be developed based on the strengths of these two models for better pothole identification and classification.

This thesis has also made a significant contribution towards the development of a standard data set for pothole detection. In order to further facilitate the data collection process, this thesis has proposed a cloud-based architecture to report road potholes and share information on road potholes with various stakeholder. This will help road users to report portholes automatically. Once the data have been processed, the database of potholes will be updated in real-time. The authorities responsible for maintaining road condition can access the road pothole database to prioritise their maintenance work. The companies that provide route planning services can use the road potholes database to optimise the routes. The route with more potholes will take a long time to commute but may also cause damage to the vehicle or serious injuries. This solution is also able to improve the times and safety of public travel.

6.4 Limitation

The limitations encountered in this research are due to hardware and data set. The scientific approach has been taken to overcome those limitations.

Hardware Limitation: A general iOS smartphone was used to collect imagery and sensory data. Due to hardware limitation, the GPS sampling rate was set at 1 HZ. The accelerometer data sampling rate is limited to 100 Hz on the latest model on an iOS smartphone. These smartphones can sample at 4000 Hz, but the limitation has been imposed to save battery life. A vehicle running at 30 miles per hour covers over 13 meters every second. 1 Hz or 100 Hz is not an optimal sampling rate to effectively detect road potholes. The sensory data sampling rate was set at 100 Hz. However, few samples had a lower sampling frequency than 100 Hz. These samples were resampled, as discussed in section 4.3.1.1. There are dependencies on the orientation of the vehicle's smartphone and speed, as discussed in section 4.2.2 and section 4.2.3.

Dataset size and variation: The data set used in this study was collected using only one vehicle. The data set will have more variations when data are collected using various vehicles as suspension on each vehicle varies. To compensate for this, scaling +- 5% was applied (4.3.1.3). This will be compensated as this study proposed data sharing from the users. Later in the time, the dataset will have data from across the world as the proposed architecture is a cloud-based crowdsourcing method.

6.5 Future Research

The approach uses a standard smartphone and a generic application to collect data. This study was conducted offline. However, given the method's advantages, the data can be collected and processed in real-time. As proposed in this research, images and accelerometer data collected from different vehicles can be uploaded to a cloud server in real-time. Later, the data can be downloaded from the server to do processing and classification using the proposed model. The classified images need to be tagged with GPS data, and notification could be sent to the users who have subscribed for the same.

1. Future research can include improving the database with more images /videos and sensory data, and data diversity. The future work can explore optimization algorithms to determine optimal parameters for the model and refine the proposed Deep Learning model to improve accuracy.
2. Future research could utilize more advanced Deep Learning methods such as Fast R-Convolution Neural Networks to process the data and produce more accurate classification results. As IoT devices with more and more computational power become available to the public, it will become possible to deploy advanced techniques that require high hardware and software requirements.

3. Future research could consider using modern communication protocols and network technologies to communicate the proposed method's results in real-time to road infrastructure such as digital road signs. These signs could warn the drivers of potholes that may cause damages to their vehicles and potential hazards to their physical health.

REFERENCES

Bouilloud, L., Martin, E., Habets, F. & Boone, A., 2009. Road Surface Condition Forecasting in France. *American Meteorology Society.*

Chen, Y. L., Mohammad , R., Jahanshahi, R. & Preetham , M., 2016. Inexpensive Multimodal Sensor Fusion System for Autonomous Data Acquisition of Road Surface Conditions. *IEEE Sensors Journal,* pp. 7731-7743.

Gueta, L. B. & Sato, A., 2017. *Classifying Road Surface Conditions Using Vibration Signals.* Mallaysia, IEEE Explore.

Islam , M. & Sadi, . M. S., 2018. *Path Hole Detection to Assist the Visually Impaired People in Navigation.* Dhaka, Bangladesh, IEEE Explore, pp. 268-273.

Ahmed , A. M. A.-S. & Tao, H., 2017. Review of Deep Convolution Neural Network in Image Classification. *International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications.*

Alqudah , Y. & Sababha, B. H., 2017. *Road Surface Conditions Using Embedded Smartphone Sensors.* Jordan, IEEE Xplore, pp. 177-181.

Artis, M., Girts, S. & Reinholds, Z., 2011. Real Time Pothole Detection using AndroidSmartphones with Accelerometers. *2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS),* pp. 1-6.

Astarita, V. & Vittoria, M., 2012. A mobile application for road surface quality control:UNIquALroad. *Procedia Social and Behavioural Sceince,* Volume 34.

Banica, C., Paturca, S. V. & Grigorescu, S. D., 2017. *Data Acquisition and Image Processing System for Surface Inspection.* ROmania, IEEE Xplore, pp. 28-33.

Basavaraju, A., Du, J., Fujie, Z. & Jim, J., 2020. A Machine Learning Approach to Road Surface Anomaly Assessment Using Smartphone Sensors. *IEEE SENSORS JOURNAL,* 20(5).

BBCNews, 2019. *BBC News.* [Online].

Burke, J. A. E. & Srivastava, M. B., 2006. Participatory sensing. *UCLA: Center for Embedded Network Sensing.*

Buza, E. & Omanovic, S., 2013. A. Pothole detection with image processing and spectral clustering.. *In Proceedings of the 2nd International Conference on Information Technology and Computer Networks, Athens, Greece,* pp. 48-53.

Chang, K., Chang, J. & Liu, J., 2005. Detection of pavement distresses using 3D laser scanning technology. *Computing in Civil Engineering,* pp. 1-11.

Chao , W., Zhen , W., Simon , H. & Marc , S., 2020. An Automated Machine-Learning Approach for Road Pothole Detection Using Smartphone Sensor Data. *Sensors,* Volume 20.

Chao, Zhen & Simon, 2020. An Automated Machine-Learning Approach for Road Pothole Detection Using Smartphone Sensor Data. *Sensors.*

Chellaswamy, C., Saravanan, M. & Kanchana, E., 2020. *Deep Learning Based Pothole Detection and Reporting System. I.* s.l., 7th International Conference on Smart Structures and Systems (ICSSS), pp. 1-6.

Chen, E. & Wang, C., 2019. Application of Improved Convolutional Neural Network in Image Classification. *2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI).*

Chen, K. & Lu, M., 2013. *CRSM: Crowdsourcing based Road Surface Monitoring.* China, IEEE Explore.

Chen, X. & Lin, X., 2014. Big data deep learning: challenges and perspectives. *IEEE access,* Issue 2, pp. 314-325.

Chen, Y. L. & et al., 2016. Inexpensive Multimodal Sensor Fusion System for Autonomous Data Acquisition of Road Surface Conditions. *IEEE Sensors Journal,* pp. 7731-7743.

Chen, Y., Zhang, P. & Wang, S., 2018. Image Feature Based Machine Learning Approach for Road Terrain Classification. *Proceedings of 2018 IEEE International Conference on Mechatronics and Automation.*

Correia, H. O. a. P. L., 2013. Automatic road crack detection and characterization. *IEEE Transactions on Intelligent Transportation Systems,* 14(1), pp. 155-168.

D. Gailius, S. J., 2007. Ice detection on a road by analyzing tire to road friction ultrasonic noise. *ULTRAGARSAS Journal, Ultrasound Institute, Kaunas, Lithuania,* 62(2).

Davide , A. & Alessandro , G., 2012. Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine. *Ambient Assisted Living and Home Care. IWAAL 2012. Lecture Notes in Computer Science,* Volume 7657.

Dey, M., Satapathy, U. & Bhanse, P., 2019. MagTrack: Detecting Road Surface Condition using Smartphone Sensors and Machine Learning. *IEEE Region 10 Conference (TENCON),* pp. 2485-2489.

Douangphachanh, V. & Oneyama, H., 2013. Estimation of road roughness condition from smartphones under realistic settings. *In Proceedings of the 2013 13th International Conference on ITS Telecommunications (ITST), Tampere, Finland,.*

Doungphachanh, V. & Oneyama, H., 2013. Estimation of Road Roughness Conditon from Smartphones under Realistic Settings. *International Conference on ITS Telecommunications.*

EL-kady, A., Emara, K. & Shaaban, E., 2019. *Road Surface Quality Detection using Smartphone Sensors: Egyptian Roads Case Study.* s.l., IEEE Explore.

Ericsson, 2016. *Ericsson Mobility Report.* [Online] Available at: https://www.iot.gen.tr/wp-content/uploads/2016/08/160307-Ericsson-mobile-report-MWC-Update-edition.pdf

F. Farnood Ahmadia, M. J. V. Z., 2008. The Application of Neural Networks , Image Processing and Cad-based Environments Facilities in Automatic Road Extraction and Vectorization from High Resolution Satellite Images. *Semantic Scholar Computer Science.*

Farabet, C., Couprie, C., Najman, L. & LeCun, Y., 2013. Learning hierarchical features for scene labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 1915-1929.

Fukushima, K., 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics,* pp. 193-202.

Fu, M., Yuan , Y. & Lu, X., 2015. *Unsupervised feature learning for scene classification of high resolution remote sensing image.* Chengdu, China, 2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP).

Gao, Q., Liu, J. & Zhang, H., 2012. Enhanced fisher discriminant criterion for image recognition. *Elsevier.*

Girshick, R., 2015. Fast R-CNN. *arXiv.org.*

Gov-Report, 2021. *Connected and Automated Vehicles: market forecast 2020.* [Online].

Graham, M. & Drobnjak, I., 2018. A supervised learning approach for diffusion MRI quality control withminimal training data. *Elseveir,* Issue 178, pp. 668-676.

Gunawan, F. E., Yanfi & Soewito, B., 2015. A Vibratory-based Method for Road Damage Classification. *2015 International Seminar on Intelligent Technology and Its Applications.*

H.OliveiraandP.L.Correia, 2014. Crackit-animageprocessing toolbox for crack detection and characterization. *Proceedings of IEEE International Conference on Image Processing,* pp. 798-802.

Hariharan, B., Arbelez, P., Girshick, R. & Malik, J., 2014. Simultaneous detection and segmentation. *Lecture Notes in Computer Science,* p. 297– 312.

Harikrishnan , P. & Gopi, V., 2017. Vehicle Vibration Signal Processing for Road Surface Monitoring. *IEEE Sensors Journal,,* pp. 5192-5197.

He, Z., 2010. *A New Feature Fusion Method for Gesture Recognition Based on 3D Accelerometer.* China, IEEE Xplore.

Hinton, G. E. et al., 2012. Improving neural networks by preventing co-adaptation of feature detectors. *Arvix.*

Huidrom, L. & Das, L., 2013. "Method for automated assessment of potholes, cracks and patches from road surface video clips,". *Procedia Soc. Behav. Sci.,* pp. 213-321.

Huijuan , W. & Jiping, . C., 2019. One-Dimensional CNN-Based Intelligent Recognition of Vibrations in Pipeline Monitoring With DAS. *Journal of Lightwave Technology,* 37(17).

Hui, L. & Yu-jie, S., 2018. Research on face recognition algorithm based on improved convolution neural network. *IEEE Xplore.*

Ikeda, Y. & Inoue, M., 2018. An Estimation of Road Surface Conditions Using Participatory Sensing. *International Conference on Electronics, Information, and Communication (ICEIC),* pp. 1-3.

Inoue, M. & Jiang, M., 2017. Motion-blur-free vidoe shooting system based on frame-by-frame intermittent tracking. *ROBOMECH.*

Ioffe, S. & Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXivLabs.*

Iphone, 2019. *Apple Iphone Xs Max.* [Online] Available at: https://support.apple.com/kb/SP780?viewlocale=en_GB&locale=en_GB

Iqbal, R, R., Maniak , T. & Karyotis, C., 2019. Intelligent remote monitoring of parking spaces using licensed and unlicensed wireless technologies. *IEEE Network,* pp. 23-29.

ItakuraY, T., 1982. Polarizationon- board sensor for detection of wet road surface, using polarization characteristics of reflection. *IEEE,* p. 66.

Jahangiri, A., Rakha, H. & Dingus, T., 2015. Adopting machine learning methods to predict red-light running violations. *Proceedings of IEEE International Conference on Intelligent Transportation Systems,* p. 650–655.

Jang, J., Smyth, A. & Yang , Y., 2015 . *Road surface condition monitoring via multiple sensor-equipped vehicles.* Hong KOng, IEEE Infocom, pp. 43-44.

Jing , L. & Tian, Y., 2020. Self-supervised Visual Feature Learning with Deep Neural Networks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 1-1.

Kaiming He, X. Z. S. R. J. S., 2014. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *arXiv:1406.4729.*

Karyotis, C. et al., 2019. *Deep Learning for Flood Forecasting and Monitoring in Urban Environments.* FL, USA, 019 18th IEEE International Conference On Machine Learning And Applications (ICMLA).

Kim, T. & Ryu, S.-K., 2014. Review and Analysis of Pothole Detection Methods. *Journal of Emerging Trends in Computing and Information Sciences,* 5(8).

Krizhevsky, A., Sutskever, I. & Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. *dvances in Neural Information Processing Systems,* p. 1097–1105.

Lecun, Y., Patrick, H., Leon, B. & Yoshua, B., 1998. Object Recognition with Gradient-Based Learning. *AT&T Shannon Lab.*

Lee, S.-M., Yoon, S. M. & Cho, H., 2017. Human activity recognition from accelerometer data using Convolutional Neural Network. *2017 IEEE International Conference on Big Data and Smart Computing (BigComp),* pp. 131-134.

Liu, X., Wang, M. & Wu, X., 2015. Visual classification by l1-hypergraph modeling, IEEE Transactions on Knowledge Data Engineering 27 (9). p. 2564– 2574.

Luo, C., Ni, B., Yan, S. & Wang, M., 2016. Image classi cation by selective regularized subspace learning. *IEEE Transactions on Multimedia,* pp. 40-50.

Mandal, V., Uong , L. & Adu-Gyamfi, Y., 2018. *Automated Road Crack Detection Using Deep Convolutional Neural Networks.* Seattle, IEEE Xplore, pp. 5212-5215.

Mednis, A., Strazdins, G. & Zviedris, R., 2011. Real time pothole detection using android smartphones with accelerometers.. *In Proceedings of the 2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS).*

Medvedev, M. & Pavlov, V., 2020. Road Surface Marking Recognition and Road Surface Quality Evaluation Using Convolution Neural Network. *2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon),* pp. 1-2.

Mohan, P., Ramjee, V. & Nericell, R., 2008. Rich monitoring of road and traffic conditions using mobile smartphones. *In Proceedings of the 6th ACM conference on Embedded network sensor systems, Raleigh, NC, USA,* pp. 357-358.

Nobis, F., Geisslinger, M., Weber, M. & Betz, J., 2019. A Deep Learning-based Radar and Camera Sensor Fusion Architecture for Object Detection. *Sensor Data Fusion: Trends, Solutions, Applications (SDF), Bonn,* pp. 1-7.

Pereira, V., Tamura, S., Hayamizu, S. & Fukai, H., 2018. A deep learning-based approach for road pothole detection in timor leste. *IEEE International Conference on Service Operations and Logistics, and Informatics ,* pp. 279-284.

Qiu, G., Du, R., Gao, K. & Liu, L., 2020. *Abnormal Road Surface Detection Based on Smart Phone Acceleration Sensor and Crowdsourcing.* Greece, IEEE, pp. 222-227.

R. Medina, J. L. E. Z. a. J. G.-G.-. B., 2014. Enhanced automatic detection of road sur- face cracks by combining 2d/3d image processing techniques. *Proceedings of IEEE International Conference on Image Processing,* pp. 778-782.

RAC, 2016. *www.racfoundation.org.* [Online] Available at: https://www.racfoundation.org/media-centre/pothole-compensation-claims-2015-16.

Raj , M. & Murugan, S. S., 2019. *Underwater Image Classification using Machine Learning Technique.* Ernakulam, India, International Symposium on Ocean Technology (SYMPOL).

Report, 2013. *Contributory factors for reported road accidents (RAS50).* [Online] Available at: https://www.gov.uk/government/statistical-data-sets/ras50-contributory-factors

Report, 2016. .. [Online].

Report, 2016. *atseuromaster.co.uk.* [Online] Available at: http://atseuromaster.co.uk/motoring-information-hub/the-motor-industry/what-is-the-cost-of-pothole-damage-to-drivers-801831923/

Roh, Y., Heo, G. & Whang, S., 2021. A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective. *EEE Transactions on Knowledge and Data Engineering,* Volume 33, pp. 1328-1347.

Rongshi, D. & Yongming, T., 2019. *Accelerator Implementation of Lenet-5 Convolution Neural Network Based on FPGA with HLS.* China, 3rd International Conference on Circuits, System and Simulation.

Roychowdhury, S., Zhao, M. & Wallin, A., 2018. Machine Learning Models for Road Surface and Friction Estimation using Front-Camera Images. *IEEE Xplore.*

S. Mathavan, K. K. a. M. R., 2015. A review of three-dimensional imaging technologies for pavement distress detection and measurements. *IEEE Transactions on Intelligent Transportation Systems,* pp. 2353-2362.

S. Ren, K. H. R. G. a. J. S., 2015. Faster R-CNN: To- wards real-time object detection with region proposal net- works. In NIPS,.

S. Zhan, Q. Q. T. X. H. L., 2015. Face detection using representation learning. *Neurocomputing,* pp. 19-26.

Sabanovic, E. & Zuraulis, V., 2020. Identification of Road-Surface Type Using Deep Neural Networks for Friction Coefficient Estimation. *Sensors.*

Sattar, S., Chapman, M. & Li, S., 2018. Road surface monitoring using smartphone sensors: A review. Sensors 2018, 18, 3845.. *Sensors.*

SciPy, 2020. *https://docs.scipy.org/doc/scipy/reference/tutorial/interpolate.html.* [Online].

Shaoqing Ren, K. H. R. G. J. S., 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv.org.*

Shen, W., Wu, G., Sun, Z. & Xiong, W., 2011. *Study on classification methods of remote sensing image based on decision tree technology,.* Nanjing, China, International Conference on Computer Science and Service System (CSSS).

Shi, Y., Cui, L., Qi, Z. & Meng, F., 2016. Automatic Road Crack Detection Using Random Structured Forests. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.*

Singh, G., Bansal, D. & Aggarwal, N., 2017. Smart patrolling: An efficient road surface monitoring using smartphone sensors and crowdsourcing. *Elsevier,* Volume 40, pp. 71-88.

Steinkraus, D., Buck, I. & Simard, P., 2006. Using GPUs for machine learning algorithms. *IEEE.*

Sun, Y., Wang, . X. & Tang, X., 2013. Deep convolutional network cascade for facial point detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.*

Tsung-Yi Lin, P. D. R. G. K. H. B. H. S. B., 2017. Feature Pyramid Networks for Object Detection. *Computer Vision and Pattern Recognition,* Volume 1.

UedaK, H., 1995. Detection of wet road surfaces by image processing. pp. 40-45.

Ukhwah, E., Yuniarno, E. & Suprapto, Y. K., 2019. Asphalt pavement pothole detection using deep learning method based on yolo neural network. *International Seminar on Intelligent Technology and Its Applications ,* pp. 35-40.

Varona, B., Monteserin, A. & Teyseyre, A., 2019. A deep learning approach to automatic road surface monitoring and pothole detection.. *Personal and Ubiquitous Computing,* pp. 1-16.

Wang, M. et al., 2015. Facilitating image search with a scalable and compact semantic mapping, IEEE Transactions on Cybernetics. pp. 1561-1574..

WavefrontLabs, 2019. *Wavefrontlabs.* [Online] Available at: http://wavefrontlabs.com

Wiratmoko, A. D. & Syauqi, A. W., 2019. *Design of Potholes Detection as Road's Feasibility Data Information Using Convolutional Neural Network (CNN).* s.l., IEEE Xplore.

Xia, W., 2018. *An Approach for Extracting Road Pavement Disease from HD Camera Videos by Deep Convolutional Networks.* Shanghai, China, s.n., pp. 418-422.

Yagi, K., 2010. xtensional smartphone probe for road bump detection. *In Proceedings of the 17th ITS World Congress (ITS Japan and ITS America ERTICO).*

Yamashita, R., Nishio, M., Do, R. K. G. & Togashi, K., 2018. Convolutional neural networks: an overview and application in radiology. *CrossMark.*

Yamashita, R. & Nisho, M., 2018. Convolutional neural networks: an overview and application in radiology. *Springer.*

Yi, C., Chuang , Y. & Nian, C., 2015. Toward Crowdsourcing-Based Road Pavement Monitoring by Mobile Sensing Technologies. *IEEE Transactions on Intelligent Transportation Systems,* pp. 1905-1917.

Zang, K. & Jie , S., 2018. Assessing and Mapping of Road Surface Roughness based on GPS and Accelerometer Sensors on Bicycle-Mounted Smartphones. *Sensors,* Volume 914.

Z. H., 2010. *A New Feature Fusion Method for Gesture Recognition Based on 3D Accelerometer.* China, IEEE Xplore.

Zhao, Y. H. a. C., 2010. A local binary pattern based meth- ods for pavement crack detection. *Journal of Pattern Recognition Research,* 5(1), pp. 140-147.

Zhou, Y., 1994. *Multi-sensor image fusion," Proceedings of 1st International Conference on Image Processing.* s.l., IEEE Xplore, pp. 193-197.

Zou, Q. et al., 2012. Crack- tree: Automatic crack detection from pavement images. *Pattern Recognition Letters,* 33(3), p. 227–238.