A Model for Detection of Safety Hazards in Construction Sites using Convolutional Neural Networks

By

Anthony Mwangi Maina

077465

Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Data Science and Analytics at Strathmore University

Institute of Mathematical Sciences and iLab Africa

Strathmore University

Nairobi, Kenya

June, 2025

Declaration and Approval

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the dissertation contains no material previously published or written by another person except where due reference is made in the dissertation itself.

Student's Name: **Anthony Mwangi Maina**

Sign: _____ Date: 23rd May 2025

Approval

The dissertation of **Anthony Mwangi Maina** was reviewed and approved by the following:

Dr. Kennedy Senagi,
Lecturer, Insititute of Mathematical Sciences,
Strathmore Univeristy.

Dr. Godfrey Madigu,
Dean, Institute of Mathematical Sciences,
Strathmore University.

Prof. Bernard Shibwabo,
Director of Graduate Studies,
Strathmore University.

Abstract

Safety should be at the forefront of every construction endeavor as it can be a matter of life or death. As such project success depends heavily on the safety protocols enacted within a site. Safety assurance, has over the past year been under the management of safety professionals. They are responsible for risk analysis, detection of safety hazards, safety protocol compliance, housekeeping etc. This however is quite an onerous task prone to errors and omissions. For example, an individual may miss to detect exposed electrical cabling, a significant safety risk to all within the working area or vicinity. The purpose of this paper is to study the utilization of deep learning techniques in detection of safety hazards on site. The hazards in consideration were fire, electrical hazards and inhalation that can be detected visually. The research workflow constituted data collection, data preparation, model training and validation. Images were collected from various construction sites in Nairobi and augmented by additional data points from the internet. The images were annotated, prepared and augmented using Roboflow, an online annotation and image-preparation tool creating a dataset of 3000 images. The images were split in the ratio of 80:10:10 and then used to train and validate three pre-trained models namely the YOLO version 8, Faster Regional-based Convolutional Neural Network (R-CNN) and Single Shot Detector Single Shot Detector (SSD) models. A mean average precision (mAP) of 0.25, 0.16 and 0.19 was achieved for the respective models, across the three classes or hazards. These results indicated the potential for use of computer vision in safety hazard detection. The You Only look Once (YOLO) version model proved superior in terms of results and was adopted for inference in a web application through a Representational State Transfer (REST) Application Programming Interface (API). This indicated possibility of use in real world setting furthermore, the model could be deployed to a CCTV camera where it can be continually trained to improve its result output. Safety professionals and stakeholders in construction projects would then be able to identify and tackle safety risks in a shorter time span enhancing their capacity to preserve life and health while still adequately managing project resources.

Keywords: Deep learning, Computer vision, Construction, Safety

Table of Contents

List of Figures

List of Tables

List of Abbreviations

API Application Programming Interface

2D Two Dimensional

ACID Atomicity Consistency Integrity Durability

AUC Area Under Curve

BIM Building Information Modeling

CRISP-DM Cross-Industry Standard Process for Data Mining

CNN Convolutional Neural Network

ERD Entity Relationship Diagram

F-CNN Fully Connected Convolutional Neural Network

FCN Fully Connected Neural Network

FN False Negative

FP False Positive

IOU Intersect Over Union

IT Information Technology

mAP Mean Average Precision

mAR Mean Average Recall

mp Megapixels

REST Representational State Transfer

R-CNN Regional-based Convolutional Neural Network

ROC Receiver Operating Characteristic

ROI Region of Interest

RPN Regional Proposal Network

SQLite Structured Query Language (Light Version)

SSD Single Shot Detector

TN True Negative

TP True Positive

UI User Interface

UML Unified Modeling Language

VGG Visual Geometry Group

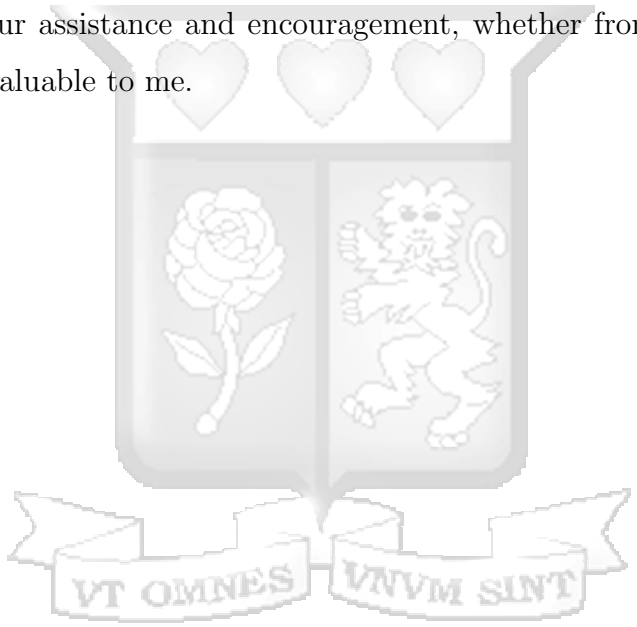
YOLO You Only look Once

Acknowledgments

I would like to express my sincere appreciation to my dissertation supervisors, Dr. Kennedy Senagi , for his invaluable guidance, support, and encouragement throughout this research. His expertise and insightful feedback have been instrumental in shaping the direction and quality of this dissertation.

I am also grateful to Strathmore University for providing the resources and conducive environment necessary for carrying out this study. Your contribution has greatly enriched the findings and analysis presented in this dissertation.

I also extend my gratitude to everyone who has played a role, big or small, in completing this dissertation. Your assistance and encouragement, whether from family, friends, or others, have been invaluable to me.



Chapter 1: Introduction

1.1 Background

Construction is one of the major industries in Kenya as it plays a vital role in the economy through the supply of housing and civil infrastructure. It accounts for almost 10% of the country's gross domestic product and employs almost 300,000 people within the sector (Nyerere, 2016). Despite this favorable metrics, the construction industry is still in its nascent stage. From the lack of standardization in quality to the scattered safety guideline compliance, there is need for improvement within the industry. Safety is one the nagging issues affecting the industry. According the Directorate of Occupational Health and Safety Services, in 2011 alone 16% of fatal accidents and 7% of non-fatal accidents occurred within a construction site. This translates to 64 fatalities per 100,000 workers. This is pretty high in comparison to the UK which experienced 0.44 fatalities per 100,000 workers in 2013. Furthermore, it has been acknowledged that 25%-40% of fatalities in the world's occupational setting are contributed to construction (ILO, 2005). Therefore, the issue of safety is still relevant across international boundaries and not only Kenya.

In many projects, construction stakeholders provide a framework to uphold health and safety in their work, however, poor documentation of accidents, hazards, near misses, policy violation and so on is poorly done. This leads to poor implementation of the health and safety incidents leading to safety incidents that have adverse adverse effects on the projects. It is noted that globally, 30-40 percent of accidents within the construction industry are fatal with accident rates showing little improvement (Muñoz-La Rivera et al., 2021). In an attempt to ensure safety is upheld in a construction site, many construction companies employ a safety manager or professional who is mobilized to site ensure mitigation measures are put in place to deal with any safety hazards. However, this presents a challenge owing to the dynamic nature of construction. It is a tall order to constantly monitor and record any unsafe acts, emerging hazards and apply mitigation measures to all these, for a single individual or team (Bal, 2001). There will always be risk of missing data or information, which can lead to inefficiencies in upholding safety within a site. The purpose of this research is to utilize artificial intelligence, more specifically computer vision, to aid in detection of safety hazards within a construction site. The idea is to build a predictive model that ingests image or video data from surveillance camera

and classifies unsafe material or safety hazards such as smoke, excessive dust and aid in logging of the detection. This will reduce the burden of the safety professional from having to continuously monitor site conditions and thereby focus his or her attention on mitigating solutions for safety promotion.

Computer vision is an interdisciplinary scientific field that considers the utilization of artificial intelligence algorithms to derive information from visual data. Its application is varies widely between different industries. It has been utilized in transportation through the enabling of automated traffic volume counts and autonomous vehicles, in marketing through head and eye tracking, real time vision for surgical procedures and within the semiconductor manufacturing industry for automation and quality inspections, (Diwan, 2015). Over the past decade, computer vision has also witnessed a rise in its application within the construction industry Xu et al. (2021),this is due to the vast amount of digital data collected from construction site through the use of tools such as surveillance cameras and digital images collected by the various professionals within the site (Suman Paneru, 2021). It has been seen that the images and videos which were primarily utilized to monitor progress within the construction site could also aid in other dimensions such quality and safety checks.

1.2 Problem Statement

Construction sites tend to be varied and at times unpredictable in nature hence posing a significant safety risk to the stakeholders involved in it implementation (Guo et al., 2021). It is therefore necessary to provide measures to mitigate this risk. Safety risk mitigation in construction is generally handled by safety personnel who are normally based on site. These individuals are responsible for risk assessment, hazard detection, safety protocol compliance, safety conditional surveys and recommendation of mitigation measurements. This tends to be arduous task, especially with large scale projects leading to numerous errors of omission (Zhou et al., 2015). Such omissions may include inability to effectively detect safety hazards, inadequate risk assessment, recommendation based on incomplete conditional survey data etc..

1.3 Research Aim

The primary focus of this study was to study the utility of convolutional neural networks in detection of safety hazards within a construction site in Nairobi Metropolitan Area.

1.4 Research Objectives

This research aimed to address the following objectives:

(a) To assess the use of deep learning computer vision in the context of hazards in construction.

(b) To gather the necessary image data relating to safety hazards in construction sites

(c) To train and validate different deep learning computer vision models on safety hazard images in construction sites and evaluate their performance

(d) To deploy the trained computer vision model in a cloud environment.

1.5 Research Questions

The research questions addressed in this study were as follows:

(a) What studies have been done relating to utilization of computer vision models in construction safety?

(b) Where can image data relating to construction related safety hazards be obtained for training and validation of computer vision models?

(c) How can different computer vision models be trained to detect safety hazards likely to be found on a construction site?

(d) How can a model be deployed in cloud environment?

1.6 Scope and Limitations of the Study

(a) Owing to the many architectural configurations that could be utilized for convolutional neural networks, four configurations will be chosen based on past performance.

(b) Transfer learning will be utilized for these models to shorten data training periods. Regarding the features that shall be learned and inferred, safety hazards in con-

struction will be the primary focus of research, hence other objects within an image will not be considered.

(c) The safety hazards to be studied are fire hazards, dust hazards, electrical hazards and visual smoke detection.

1.7 Research Justification

Safety is a core component to every construction endeavor, and if not properly taken into account, could have devastating effects on the project, affecting not only project program but the exposing the stakeholders to significant legal risk. Furthermore, the reputation of the stakeholders is jeopardized if they are perceived to be nonchalant about it. The cost of the projects shoots up, with the overall construction cost affected most as opposed to costs associated with preliminaries and prime materials used for construction. The profitability of the construction endeavor is greatly reduced creating financial risk for contractors and consultants associated with construction.

Chapter 2: Literature Review

2.1 Computer Vision

Computer vision is the sub-domain of deep learning within the machine learning space. It involves the processing of visual data from image and video-format content through the utilizing learning algorithms in artificial intelligence. This is a relatively new space in the field of computer science and evolved because of the proliferation of data and the advancement in computational processing power. According to Voulodimos et al. (2018), deep learning is a rich family of methods encompassing neural networks, hierarchical probabilistic models, unsupervised and supervised feature algorithms. These techniques have improved over the years with the first computer vision model, AlexNET configured in 2012 by (Krizhevsky et al., 2012). Convolutional neural networks aim to mimic the visual cortex of the brain and interpret images as humans do. They consist of three main layers, the convolutional, pooling and fully connected neural network layers. An example of this network can be shown in Figure 2.1

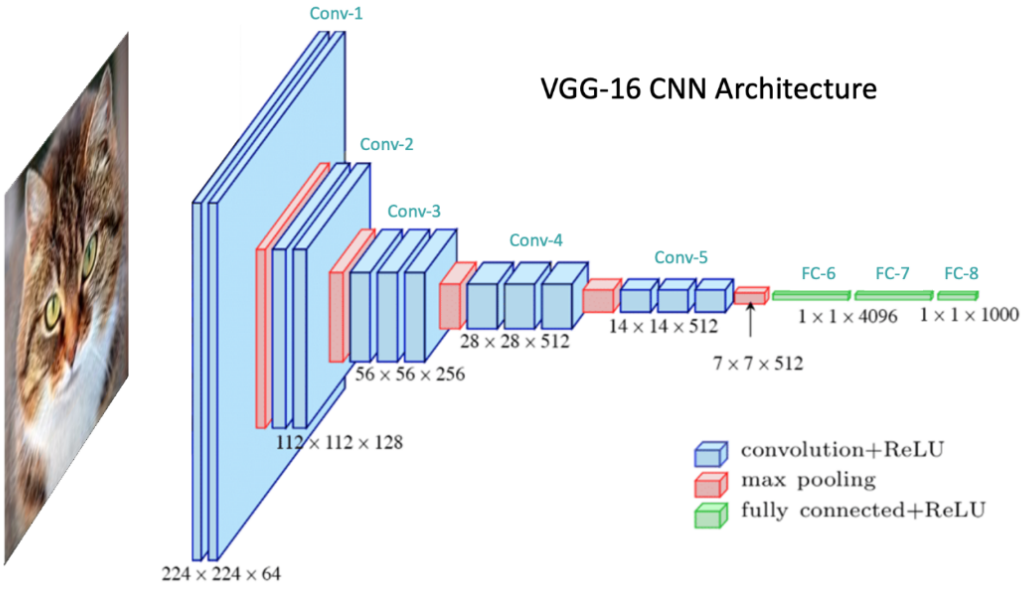


Figure 2.1: Image representation of the standard architecture of a Convolutional Neural Network, (source: openCV, 2023)

Voulodimos et al. (2018) describes the role of each of these layers. The convolutional layer consists of kernels that extracts features and important information from the raw image input through the use of training parameters. The pooling layers reduce the spatial dimensions of the image data and aid in dimensionality reduction. The fully connected neural

network layer is a weighted multi-layer perceptron with non-linear activation functions that enable image classification. The entire architecture enables application of computer vision in image classification. Other conventional applications of computer vision are object detection and object segmentation. Object detection entails the identification of an image's semantic features and their corresponding locations as pioneered by (Krizhevsky et al., 2017).

There have been advancements in object detection with modes such as Region-based convolutional neural network (R-CNN) model developed by (Girshick et al., 2014). This model was combined with a selective search, which was able to achieve a 31.4% mean Average Precision (mAP) score using the 2013 ImageNet database. Faster R-CNN, developed by Ren et al. (2017) followed suit and identified as state of the art for its accuracy of 70% on the PASCAL Visual Object Classes (VOC)4 2007. Image semantic segmentation on the other hand is a form of object detection where each pixel is classified into a fixed set of categories without differentiating an object's instance (Sun, 2016). The output is in the form of mask instead of bounding boxes in the case of image detection.

2.2 Application of Computer vision in Construction

Xu et al. (2020) conducted a critical review of literature surrounding the application of computer vision. Xu et al. (2020) introduced a fresh perspective where he viewed computer vision to operate at three levels, namely early vision, mid-level vision and high-level vision. Early vision entails feature extraction, part of the convolutional process, while mid-level vision involves image classification, object detection and object tracking. Finally, high level vision involves a higher order interpretation of the computer vision inference.

Delving deeper into computer vision application within the realm of construction, Seo et al. (2015) mapped its role in construction. This role was divided into scene-based risk identification, location-based risk identification and action-based risk identification. It is within these categories that the level of vision is defined. For example, the detection of PPE absence is a mid-level vision within the scene-based risk identification. Fang et al. (2019) conducted research involving use of convolutional neural networks to mitigate against falling from height, a scene-based risk identification. A mask-RNN was adopted

to detect construction workers use of fall protection equipment when working at height. The model had a recall of 90% and precision of 75%. Kim et al. (2016) also integrated computer vision with a fuzzy inference to monitor and assess the safety of people performing their tasks in the vicinity of plant. All are an example of high-level computer vision. Fang et al. (2020), conducted an exploratory review of use of computer vision in behavior-based safety implementation in construction, a action based risk identification application. The study was simple and qualitative, indicating that unsafe behaviors could be detected though pose estimation. The movement of workers on site are detected visually and classified as either safe or dangerous. This research poses a challenge where debate may arise on what human position is safe, furthermore certain attributes of an action may be difficult to infer. A good example would be the hoisting speed of a crane and what speed can be deemed dangerous. Luo et al. (2020) took a different approach, a full body pose estimation of construction equipment was carried out to classify the safety levels of a scene based on the relative position of construction equipment component(for example, excavation boom) to surrounding construction workers. High mean average accuracy from Stacked Hourglass and Cascaded Pyramid Network model of about 91% was attained. This was very promising research due to the severe of injuries sustained from the impact from construction equipment. However, this research was limited to equipment such as cranes and excavators that have movable, jointed parts. vehicles such as truck can also pose a safety risk even with the limited maneuverability. Majority of literature reviewed proposed the deployment of computer vision model onto onsite CCTV cameras to aid in data collection for training, validation and detection. Nguyen et al. (2018) however conducted a review of the use of computer vision for power-line inspection. In this review, it was shown that unmanned aerial vehicles and helicopters could be utilized for inspections at height. This may be helpful in safety hazard detection, especially for those working at height.

2.3 Research Gap

It was noted that not much research has done on detection of safety hazards on site, specifically static hazards. As per the author's knowledge, most research centers around the construction worker, equipment and his immediate environment. This present a research gap in the utilization of computer vision techniques in safety hazard detection

within a construction site environment. The aim of this research is to study the utilization of computer vision through convolutional neural networks in aiding safety professionals in hazard detection. These hazards will entail fire hazards, electrical hazards, visual smoke and dust detection. These hazards do not fully encompass the myriad of dangers that plague construction site but were chosen owing to their ease in visual detection.

Chapter 3: Methodology

The study followed the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, a widely adopted framework for data mining projects, shown in Figure 3.1. CRISP-DM comprises several essential phases organized in an iterative process. A detailed framework for our research methodology can be seen in Figure 3.5.

1. **Business Understanding:** In this stage, the problem statement is defined. The problem statement may in the form of business goals or research endeavors. Stakeholder engagement is carried out to ensure alignment of objectives and expectations.

2. **Data Understanding:** In this stage, data is collected and analyzed for utility and relevance to the research objectives. This enables in-depth understanding of the data requirements.

3. **Data Preparation:** In this phase, data cleaning, augmentation and feature engineering is carried out. In this stage, engagement with domain experts is critical to ensure the data is as exhaustive as possible.

4. **Modeling:** In this stage, models are scouted for, with preference for utmost diversity of the models. The models are then trained for prediction purposes.

5. **Evaluation:** The trained models are evaluated for accuracy, class distinction and other metrics relevant to the business or research objectives.

6. **Deployment:** The optimized model is embedded in business process intended as per the objectives defined. This may be in the form of a web app or dashboard. It is in this stage that the model can be employed for its purpose, however constant monitoring is done to ensure efficacy.

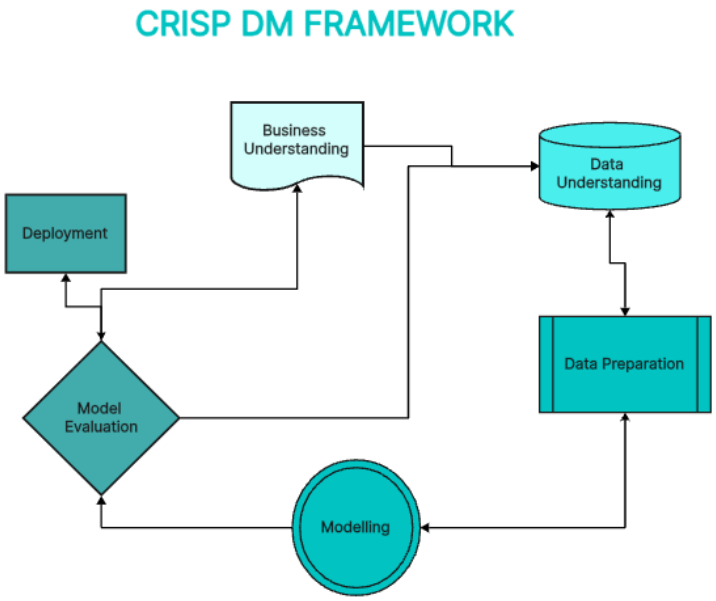


Figure 3.1: CRISP-DM framework

3.1 Business Understanding

The business understanding stage of the research encompassed the following sub-stages:

(i) Identification of key stakeholders

(ii) Outlining of research objectives

(iii) Identification of research requirements

3.1.1 Identification of Stakeholders

Safety is a core component to every construction endeavor, and if not properly taken into account, could have devastating effects on the project, affecting not only project program but the exposing the stakeholders to significant legal risk. Furthermore, the reputation of the stakeholders is jeopardized if they are perceived to be nonchalant about it. The cost of the projects shoots up, with the overall construction cost affected most as opposed to costs associated with preliminaries and prime materials used for construction. The profitability of the construction endeavor is greatly reduced creating financial risk for contractors and consultants associated with construction.

In many projects, construction stakeholders provide a framework to uphold health and safety in their work, however, poor documentation of accidents, hazards, near misses,

policy violation and so on is poorly done. This leads to poor implementation of the health and safety incidents leading to safety incidents that have adverse adverse effects on the projects. It is noted that globally, 30-40 percent of accidents within the construction industry are fatal with accident rates showing little improvement (Muñoz-La Rivera et al., 2021).

In an attempt to ensure safety is upheld in a construction site, many construction companies employ a safety manager or professional who is mobilized to site ensure mitigation measures are put in place to deal with any safety hazards. However, this presents a challenge owing to the dynamic nature of construction. It is a tall order to constantly monitor and record any unsafe acts, emerging hazards and apply mitigation measures to all these, for a single individual or team. Bal (2001) there will always be risk of missing data or information, which can lead to inefficiencies in upholding safety within a site. The purpose of this research is to utilize artificial intelligence, more specifically computer vision, to aid in detection of safety hazards within a construction site. The idea is to build a predictive model that ingests image or video data from surveillance camera and classifies unsafe material or safety hazards such as smoke, excessive dust and aid in logging of the detection. This will reduce the burden of the safety professional from having to continuously monitor site conditions and thereby focus his or her attention on mitigating solutions for safety promotion.

3.1.2 Outlining of research objectives

Clear research objectives were defined to guide the research. These objectives have been outlined in Chapter 1, Section 1.3

3.1.3 Identification of research requirements

Research requirements were guided by the research objectives. The system and model architectures necessary to solve the research need were identified and established before implementation.

Overall, the business understanding stage aided in contextualizing the research. The problem domain, stakeholder needs, and requirements to implement a computer vision solution to safety hazards in construction sites.

3.2 Data Sources and Understanding

The primary data type for this research was 2-dimensional Two Dimensional (2D) images. The images were required to be of high quality to enable the model learn the necessary patterns. Images relating to safety hazards within the construction context were collected from several construction sites in Kiambu County. We sough informed consent before data collection. With the aid of a safety professional, image data was collected relating to fire, inhalation and electrical hazards. The sites were surveyed thoroughly by the safety professional and the data was collected over a period of months. a smart phone was utilized for image collection owing to its ubiquity and reliability. However, due to the varied nature of construction sites, some hazards were identified in areas of low lighting. As such different angles and camera setting adjustments were necessary to enable quality image collection. The image data was however, not enough to fully train the intended deep learning models, hence images from the internet collected. Repositories such as Dataset Ninja proved helpful in retrieval of the images. Google images were also incorporated into the dataset. This was due to their varied nature which was beneficial to our machine learning model.

The dataset had 3000 images in total, with 1000 images per class. This was however after data augmentation, the original data size was 846 images, with 362 images related to electrical hazards, 447 images related to fire hazards and 466 images related to inhalation hazards. The data was collected from March 2025 to April 2025 over a period of 5 weeks.

Dataset link: [`https://huggingface.co/datasets/MainaEng/Safety_Hazards/tree/main`]

3.3 Data Preparation

The data collected from different sources, was dirty and noisy and this was likely to adversely affect the performance of the machine learning model. Therefore, the image data required to undergo some level of augmentation to aid in variance reduction and enhance accurate model inference within different contexts. This was especially important to enable sustainable use of the model as it was important to the research endeavor for the solution to perform well in dynamic environments.

3.3.1 Image cleaning

Xu et al. (2021) described various techniques that could be used for data pre-processing which we considered and adopted into our research. Image cropping and grey-scaling of images was done using openCV (version 4.11.0) in Roboflow to help the model focus and classify the safety hazards areas of limited lighting. It was expected the model would be embedded to and edge device such as surveillance cameras feed which switch to grey-scale in low light. Image noise reduction proved helpful in removing noise from dirty images while contrast enhancement aided in delineating different features within an image.

3.3.2 Image Annotation

The images, once collected and cleaned, were uploaded to RoboFlow Dwyer, B., Nelson, J., Hansen, T., et. al. (2024). Roboflow is an online annotation tool that aids in the creation of bounding boxes and labels for image data intended for training of machine learning models. Roboflow employs albumentations library to effect this. This tool was used to create classes; fire, electrical and inhalation hazards, after which the bounding boxes were drawn manually around the item or hazard displayed in the image. The dimensions and location of the bounding box was automatically recorded. This shortened the annotation work flow in comparison to other libraries where the bounding box properties are hard-coded and care is ensured that each bounding box is aligned to the correct image. On completion of annotation, using Roboflow, the images were resized to a standard dimension of 640 by 640 pixels for standardization of image data for all models. Examples of these images are shown in Figure 3.2, Figure 3.3 and Figure 3.4

Figure 3.2: Fire hazard image used for training, hazard enclosed by yellow bounding box

Figure 3.3: Electrical hazard image used for training, hazard enclosed by red bounding box

Figure 3.4: Inhalation hazard image used for training, hazard enclosed by purple bounding box

3.3.3 Image Augmentation

Through the use of openCV (version 4.11.0) library in Roboflow, data augmentation was then carried out to inflate the size of the data set. The images went through the following augmentations; rotation, contrast enhancement, noise addition and flipping. This quadrupled the size of the dataset from 874 to 3000 images. The data and corresponding labels were exported in YOLO V8 and COCO format for training and testing. Class balancing was monitored and maintained from the data collection stage.

3.4 Machine Learning Model

3.4.1 Data Splitting

The data (3000 images) were split into training, validation and testing data in the ratio of 80:10:10 using Roboflow online annotation tool. The splitting of the data set aided in provision of training and validation data to aid in training and optimization of the model. Test data, which comprised of unseen data was then utilized to evaluate optimized model. This inference informed final model performance.

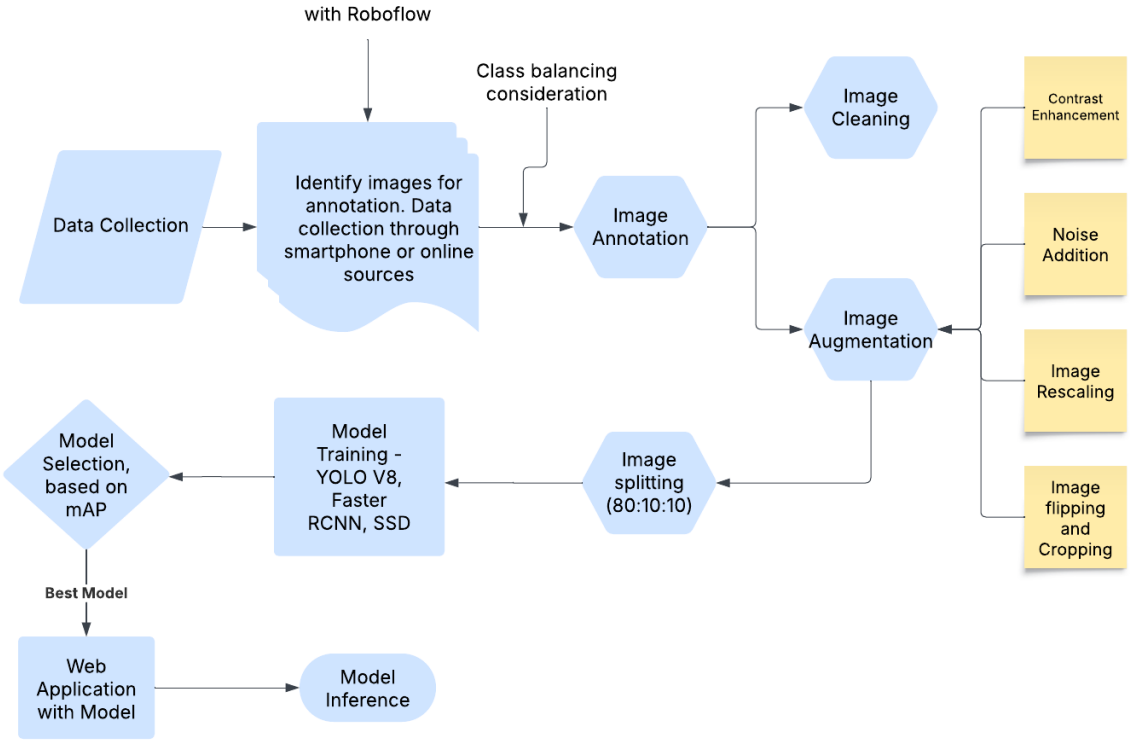


Figure 3.5: Workflow methodology for the project showing the data collection, preparation, modeling and inference

3.4.2 Model definition and Architecture

This research employed three different models. These models were pre-trained and state-of-the-art with high accuracies observed when fitted to their original datasets. The models were implemented in Pytorch (version 2.0) and Ultralytics (version 8.3.88) frameworks. Overall all code was done in python. These models were as follows:

(i) YOLO version 8

(ii) Faster R-CNN

(iii) Single Shot Detector (SSD)

3.4.3 Faster R CNN

This study implemented the Faster R-CNN model in fitting the image data. The model was obtained from the Pytorch library and loaded to the Jupyter notebook for training and inference. Faster R-CNN is a complex Convolutional Neural Network (CNN) model architecture, shown in Figure 3.6, was devised by Ren et al. (2017) as an improvement

to the fast R-CNN. It is a detection algorithm that works on the principle of regional proposal networks. In the architecture, the model proposes potential localities of the object in from the feature maps derived from a backbone CNN shared by the Regional Proposal network Regional Proposal Network (RPN) and Fast CNN. These proposed localities are known as anchor boxes derived in rectangular format. They are produced at different scales and aspect ratios to enable the algorithm detect object of different shapes and sizes.

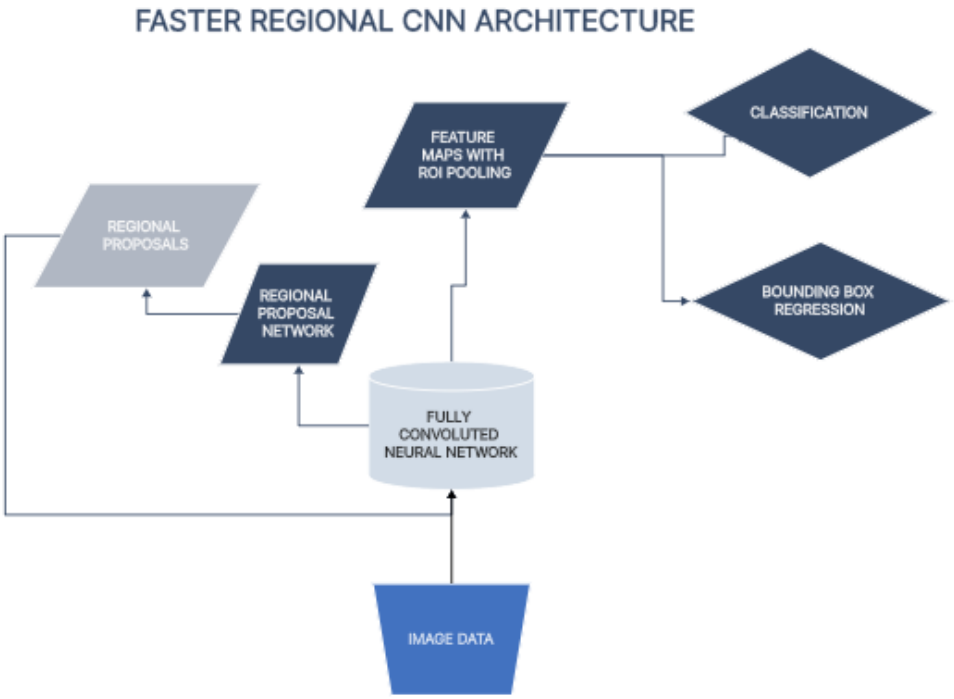


Figure 3.6: Faster R-CNN Architecture

The RPN operates through a sliding window mechanism where a small kernel (3x3) slides across feature maps from the backbone CNN. The sliding window generates an objectness score across the each part of the feature maps. The objectness score is a binary classification that denotes the presence or absence of an object in a particular region. This score also indicates the confidence of the algorithm that an object is within a certain locality. During the anchor box generation, many boxes are bound to be produced, as such, IOU score are utilized to reduce this number and maximize overlap of the region proposal and ground truth bounding box. The concept of IOU is describe in the following sections within this chapter. The region with the highest confidence score is then chosen through non-max suppression technique. Each anchor box is then furnished with the box

coordinates and objectness/confidence scores (Ren et al., 2017).

Upon generation of the regional proposals, the anchor boxes are then presented to a Fast R-CNN detector (similar to the native Fast R-CNN detector). These boxes contain feature elements which are then pooled through a process called Region of Interest Region of Interest (ROI) pooling. This process is similar to max pooling where maximal feature values are derived from anchor box feature maps. This enables creation of fixed size feature maps from the varied feature maps created from the regional proposal network. The fixed-sized feature maps are recycled to the CNN backbone where further feature extraction is done. High level features as a spatial details as well as low level features such lines and edges are retained (Ren et al., 2017).

The recycled features are then passed to a fully connected neural network (Fully Connected Neural Network (FCN)). The FCN carries out object classification and bounding-box regression. Potential bounding boxes are predicted each region with corresponding confidence scores. Optimization is accomplished through Multi-task Loss function that combines classification and regression losses by the detector (Ren et al., 2017).

3.4.4 Single Shot Detection (SSD)

Inspired by Liu et al. (2016), the Single Shot Multibox Detector is a state of the art model governed by greater detection speed. It takes a different approach from the Faster R-CNN that relies on regional proposals and resampling of the features to a backbone network. The SSD is a feed forward Fully Connected Convolutional Neural Network (F-CNN) that creates a standardized collection of bounding boxes with corresponding confidence scores. These bounding boxes are referred to as default bounding boxes and are generated at different aspect ratios and scales. This feature, similar to Faster R-CNN enables flexibility in detection of objects of different shapes and sizes. the default bounding boxes are created for all feature maps that are taken through a fully connected CNN. The architecture diagram is shown in Figure 3.7.

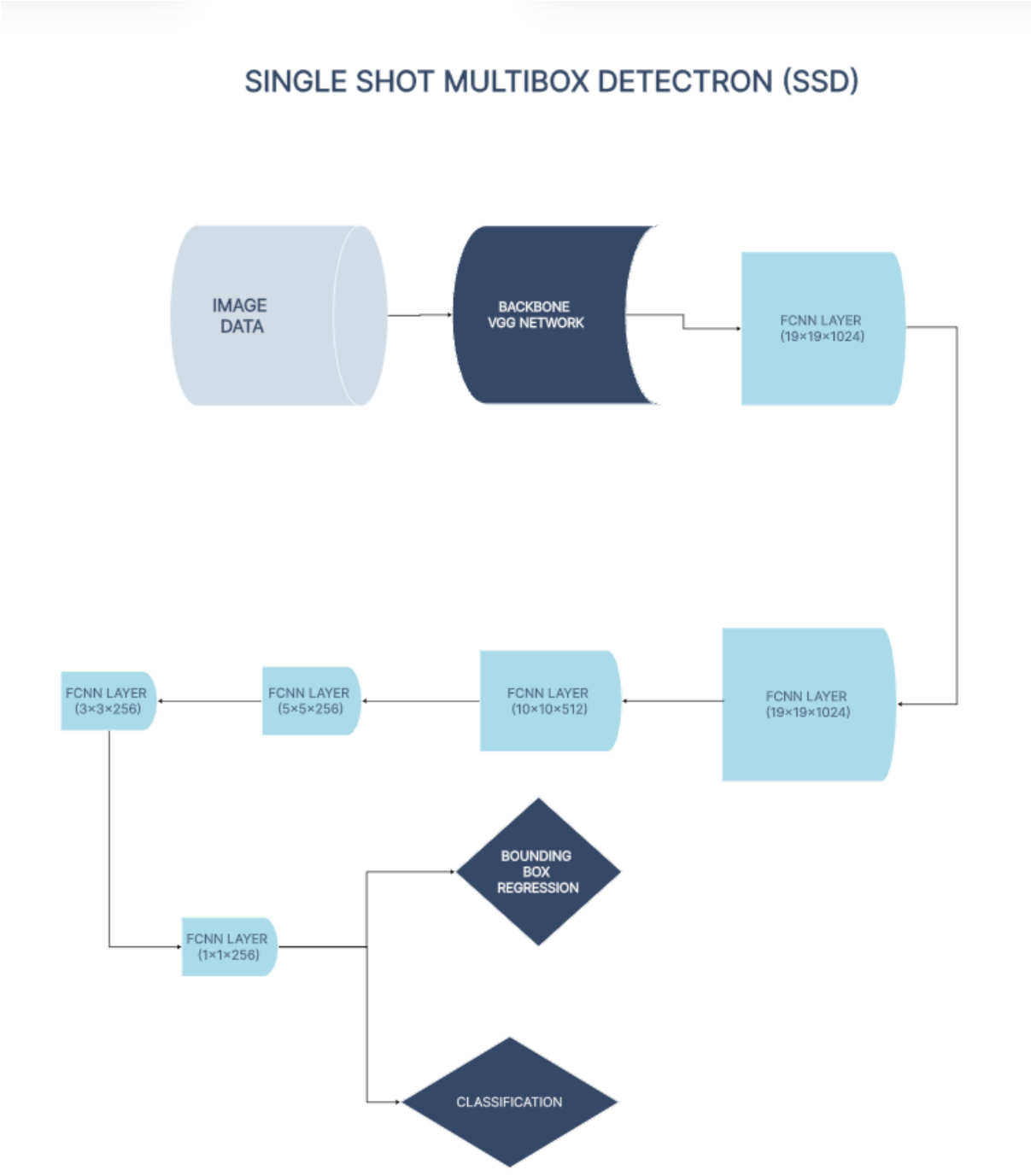


Figure 3.7: Single Shot Detector Architecture

The architecture of the SSD consists of a backbone fully connected F-CNN (typically a pre-trained Visual Geometry Group (VGG)-16) where additional fully connected layers are added to the network, size in number to be precise. The additional layers aid enhance capability of the model for multi-scale feature mapping. Images, once presented to the model, are transformed to feature maps, after which predictions are made.

3.4.5 YOLO V8

This study implemented the YOLO version 8 pre-trained computer vision model with its default hyper-parameters. This model was chosen in order to reduce the training time and data requirements necessary for deep learning (computer vision models). YOLO V8 model is a single-shot convolutional neural network that is known for its quick detection relative other object detection models. The convolutional neural network is the recommended machine learning model for learning data patterns embedded in images. The convolutional neural network is composed of three segments, the kernel layer, pooling layer and the fully connected neural network. Consequent to data ingestion, the kernel within the convolutional neural layer derives features from the image data using weights and biases that are constantly updated with data from new images.The output is a series of feature maps that comprise the essential elements of the images such as lines and edges. The pooling layer receives the feature maps and compresses the data reducing the data dimensionality. This reduces the number of parameters necessary for training simplifying the model hence reducing computational requirements necessary for training. The pooled data is then fed to a fully connected neural network that will act as the classifier. In this research, image data not only comprised the image features but also the coordinates of the bounding box around the feature (safety hazard), with the label data comprising the true bounding box coordinates and labels stored separately. The predicted labels and bounding box coordinates were then be evaluated against the true data. The architecture diagram is shown in Figure 3.8.

3.5 Model Evaluation Metrics

The major evaluation metrics considered were the intersection over union ratios and mean average precision (mAP). In order to understand these concepts, the following definitions needed to be outlined.

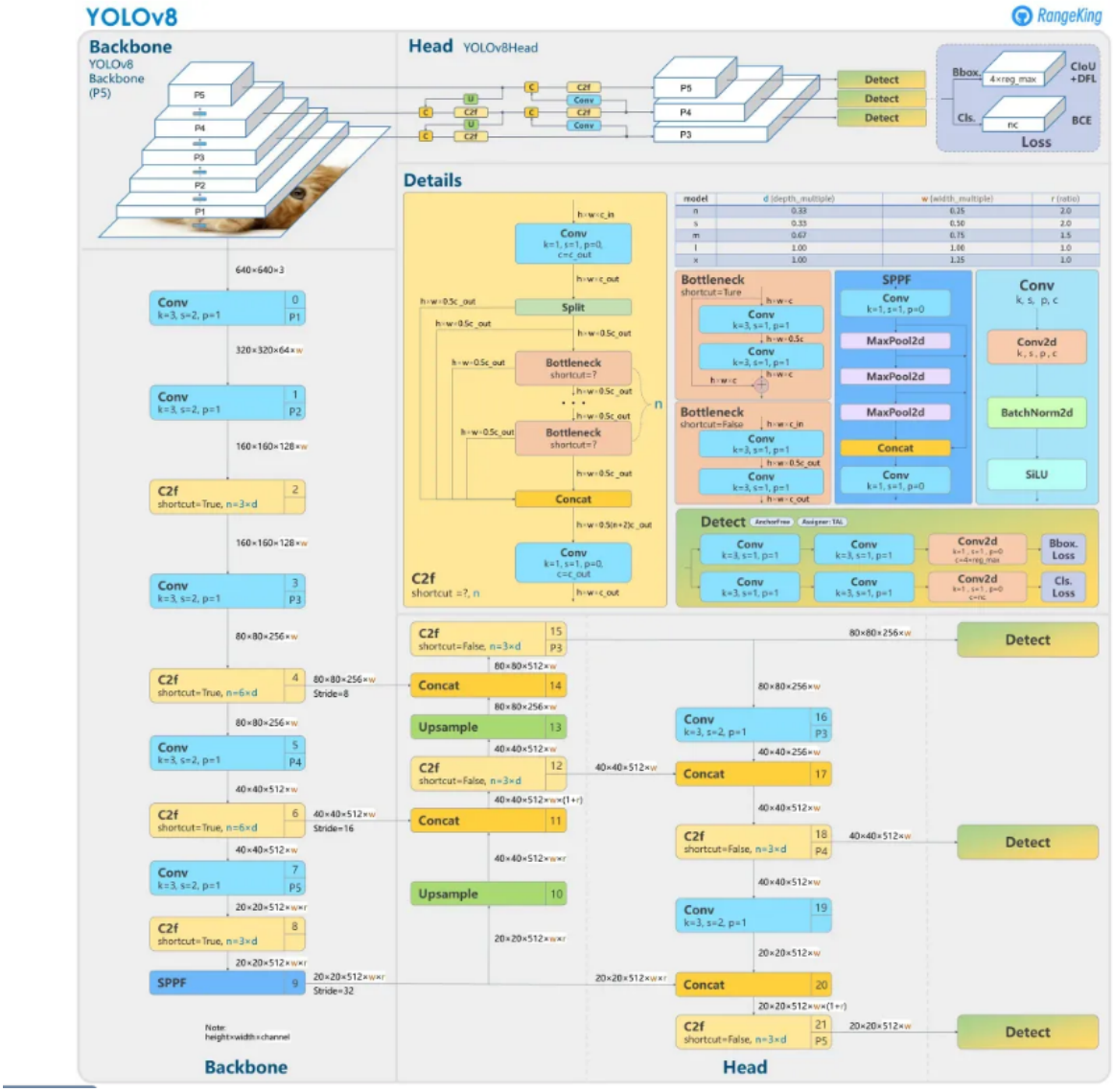


Figure 3.8: The complete architecture of the YOLO version 8 model, a series of convolutional neural layers, pooling layers and vector bottle-necking is shown(Source: Ghosh (2024))

True positivesTrue Positive (TP) These are the correct object predictions of ground truths.

False positive False Positive (FP) These are erroneous detection of non existent objects or the misplaced detection of an existing object.

False negativeFalse Negative (FN) These are undetected ground truths in an image.

True negativesTrue Negative (TN) These are not considered as there are an infinite number of bounding boxes that would construe a true negative.

This was followed by an establishment of a correct and incorrect detection. This was done through the intersection over union ratio.

3.5.1 Intersection over Union Ratio

The intersection over Union ratio, in Figure 3.9, is a value that denotes how well the predicted bounding box aligns with the ground truth bounding box. The overlapping area between the predicted and ground truth bounding boxes is considered where it was calculated and divided by the unified area of both bounding boxes, Padilla et al. (2020). The ratio was then compared with a predefined threshold of 0.5, where the IOU greater than the threshold, t, was considered accurate and vise versa.

$$IOU = \frac{area(B_{pred} \cap B_{gt}}{area(B_{pred} \cup B_{gt})} \tag{1}$$

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} =$$

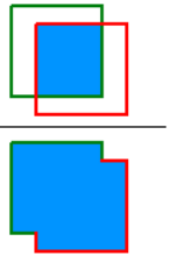


Figure 3.9: Intersection over Union (IOU)

3.5.2 Precision

Precision is the ratio of true positives to all detection (true positive plus false positives). It is a critical component in the measurement of false positives and is utilized in scenarios

where the false positives are costly to the end user. In our case, it was an important component but not critical as will be explained in chapter 6. Precision is guided by the equation below:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

The components of the equation are explained in previous parts of this section.

3.5.3 Recall

Another concept that was considered was the recall. Recall is the ability of the model to detect all relevant cases, it is the ratio of true positives to all given ground truths. The goal of the study was to train the model to obtain high precision and recall, where the number is false positive and false negatives is minimized.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

3.5.4 Area Under Curve AUC

This value was derived from the Receiver Operating Characteristic Receiver Operating Characteristic (ROC) plot. In this plot the true positive (y-axis) and false positives (x-axis) are plotted against each other. It is usually done for binary classifiers however, in our case, it was done repeatedly for the pairs of different classes and the average taken. The curve considers all data point in its plot. The plot denotes the probability that the model will detect a class in comparison to a random prediction. Therefore, an AUC of 1 is show perfect class discrimination, while an AUC of 0.5 denotes a random event. The performance of the models in relation to AUC is discussed in chapter 6.

3.5.5 F1-Score

The F1-score is a critical evaluation metric that takes into account precision and recall. It is derived from the harmonic mean of precision and recall where both metrics are balanced out to give an single figure. This figure gives a good impression of how well the machine learning model distinguishes classes provided during training. The harmonic

mean is the preferred method of calculation as it considers fractional values with different denominators. Typically the F1-score is utilized for binary classification, however in out study, the mean precision and recall across classes was used to derive the mean F1- score. Further details regarding the F1-score are provided in chapter 6 of this study with its implication on model performance.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

3.5.6 Mean Average Precision (mAP)

The mean average precision is a standard metric used for the evaluation of all object detection models. It is derived from average precision per class. It considers the recall and precision trade-off where the area under curve is calculated for each class and average across all of them. The mean average precision is provided for all epochs where the precision metrics for each class are summed up and the average taken.

$$meanAveragePrecicion(mAP) = \frac{1}{N}\sum_{i=0}^{N} AP_i \tag{5}$$

where N is the number of classes.

There was also be a temporal component to the detection as the model may be attached to a video feed. Hence the speed at which inferences were taking place was measured. This inferential time was reduced to a minimum to increase the efficacy of the model in a fast changing environment.

3.6 Model Optimization

Model optimization is the process of tuning the hyper-parameters of the model to maximize model output. It was carried out before choosing the best model for deployment. In our study, the hyper-parameters in consideration were the learning rate of the models, number of epochs and optimizer. To achieve the best results, the number of epochs were constrained to twenty to reduce computational resource utilization. Adam optimizer was considered as the optimizer on account of its good performance across all models. Learning rate of 0.01 was used with a decaying component of 0.9 to prevent the model from

being bogged down in local minima.

3.7 Deployment

The model was deployed through the Streamlit service. This tool was adopted owing to its quick prototyping capabilities and cost effective deployment. The Streamlit service provided a User Interface (UI) where one could interact with the model directly. The UI involved a simple layout where images are directly uploaded to the web application. The web application utilizes the model to generate detection of objects within the image. The object in question are explained in chapter 6.

Chapter 4: System Design and Architecture

4.1 System Modeling

The process of system modeling in this research adopted the Unified Modeling Language (UML). This is an accepted and standardized methodology in software engineering recognized for its efficacy in visualization and system architecture documentation. UML serves as a communication tool for all that act within the Information Technology (IT) industry. In our study, the use case diagram was adopted to showcase the behavioral characteristics of its components; how different part of the system interact, from actors to model development and system components (Koc et al., 2021). This diagram provides an opportunity to communicate the functionality of the system at a high level in way that is easily understandable.

Figure 4.1 illustrates the UML diagram, showcasing the various components, relationships, and behaviors within the system.

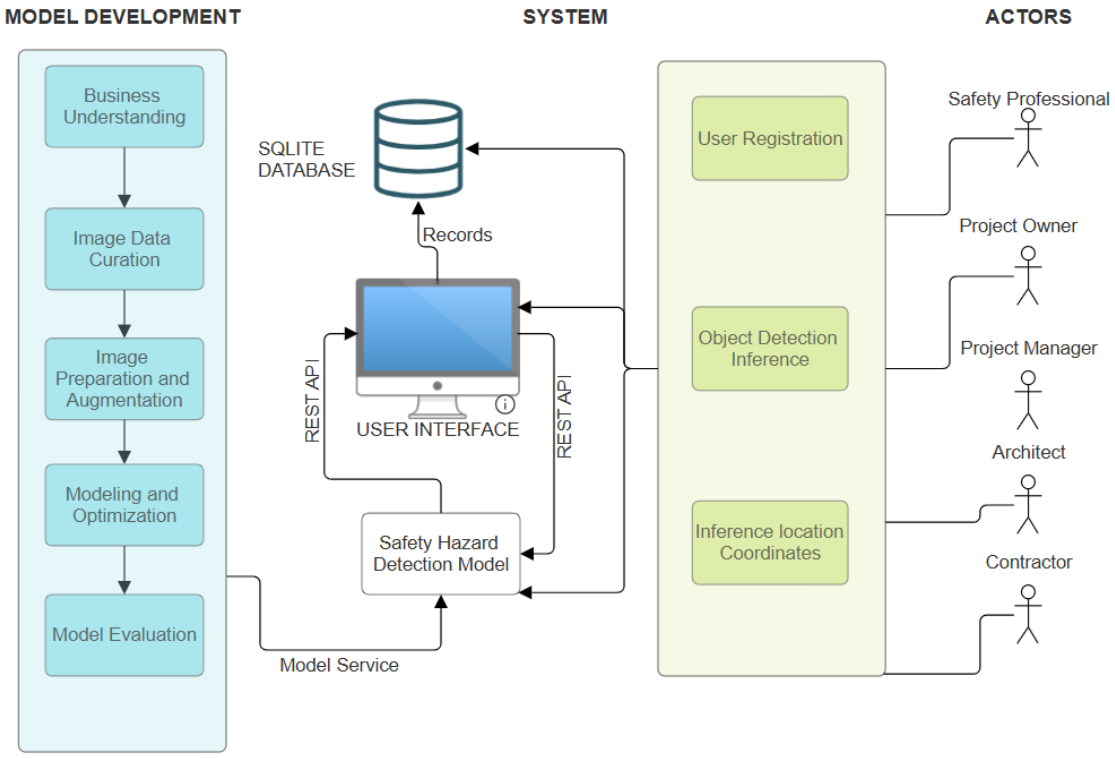


Figure 4.1: UML diagram

4.2 System Components

In our research, the main components of the system included the following: a web portal and database.

4.2.1 Database

The database architecture adopted the SQLite relational database schema, this is due to the fact that the implementation was in its early phase. In future, this architecture could be advanced to a more robust choice such as Snowflake when the data requirement are high. The SQLite schema ensured data integrity and table normalization to ensure optimal querying. In our architecture, three tables were conceptualized: detection, location and user tables as depicted in Figure 4.2. This Figure 4.2 outlined the relationships between the tables with adoption of primary and foreign keys where necessary.

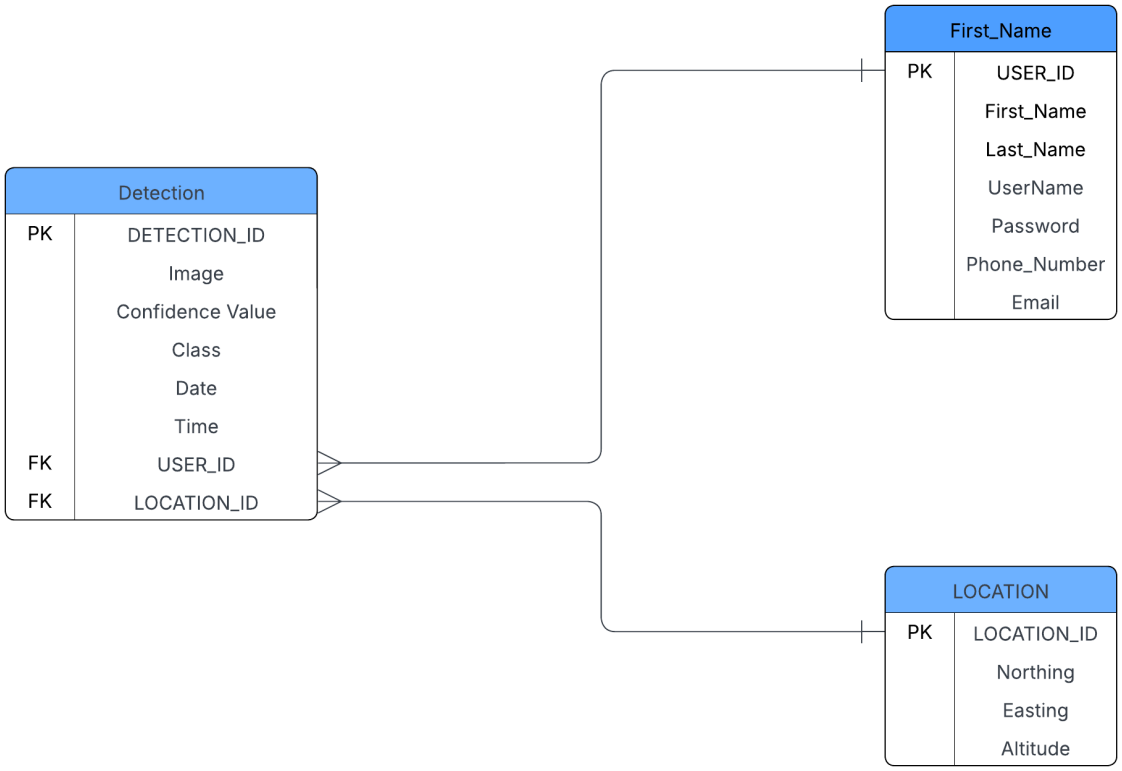


Figure 4.2: Entity relationship diagram ERD

Table 4.1 offers an elaborated breakdown of each database table that has been created.

Table 4.1: Database tables

Table	Field Name	Data Type	Description
Detections	Detection_ID	Int	Unique identifier for detections
	Image	Blob	Image to be inferred
	Class	Text	Class of object detected
	Date	Datetime	Date when detection was made
	Time	Datetime	Time when detection was made
	User_ID	Int	Foreign Key referencing User table
	Location_ID	Int	Foreign Key referencing Location Table
Location	Location_ID	int	Unique identifier for Location
	Northing	real	Northing Value
	Easting	real	Easting Value
	Altitude	float	Altitude in metres
Sector	Sector_ID	int	Unique identifier for sector
	Name	char (100)	Name of the sector
	District_ID	int	Foreign key referencing the District table
User	User_ID	int	Unique identifier for user
	First_Name	char (50)	User's first name
	Last_Name	char (50)	User's last name
	Username	text	User's application username
	Password	text	User's Password
	Phone_Number	Int)	User's phone number, including country code
	Email	Text	User's Email
	Registered_Date	DateTime	Date of user enrollment in the alerts system

4.2.2 Web Portal

The web application portal consisted the two components, namely, *Main Application Home Page* and *User Authentication and Registration Page*. The *Main Application Page* offered the platform for uploading images and predicting objects within it. The *User registration page* provided a form for signing up new users to the application as well as user authentication.

(i) Sitemap

A sitemap, shown in Figure 4.3, is a simple visualization showing the structure of the website. It outlines the various pages and shows their relationship, giving insight to website navigation for the end-user. This enhances user experience and website usability through effective website organization.

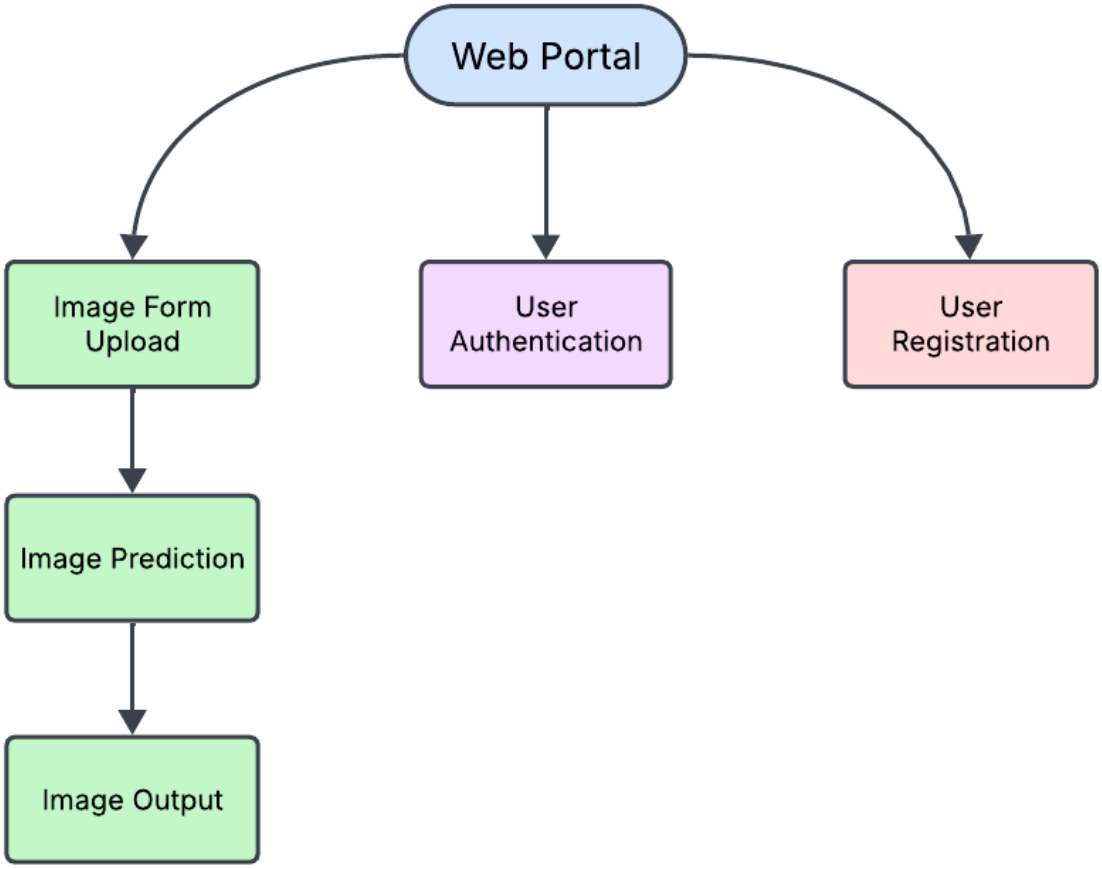


Figure 4.3: Sitemap of the web portal

(ii) Wireframes

These are schematics that represent the general layout of the website. They show the placement of different components of the website such as navigation panels, content section and other useful feature in a simplistic manner. The Lucid App [`https://www.lucidchart.com/`] was utilized in creation of the wireframes owing to its ease of use.

(a) **Main Application Page Wireframe**

The page, Figure 4.4, offers a concise overview of the image form and prediction component of the application

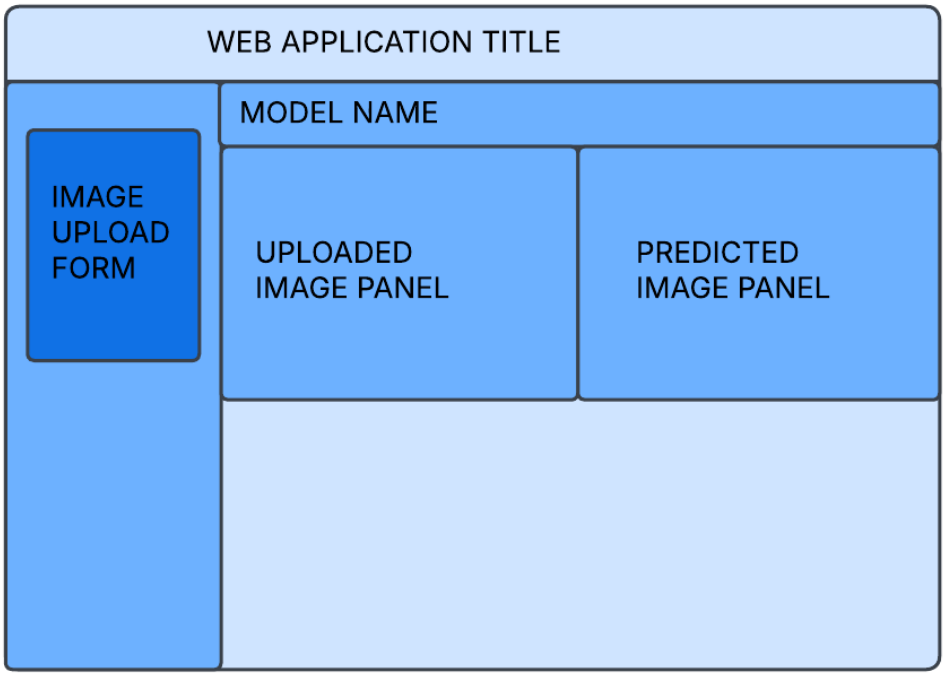


Figure 4.4: The home page wireframe

(b) **User Registration Wireframe**

Users shall have the option to sign up to the application to predict images. They shall provide basic information and and communication contacts for update communication. This is depicted in Figure 4.5

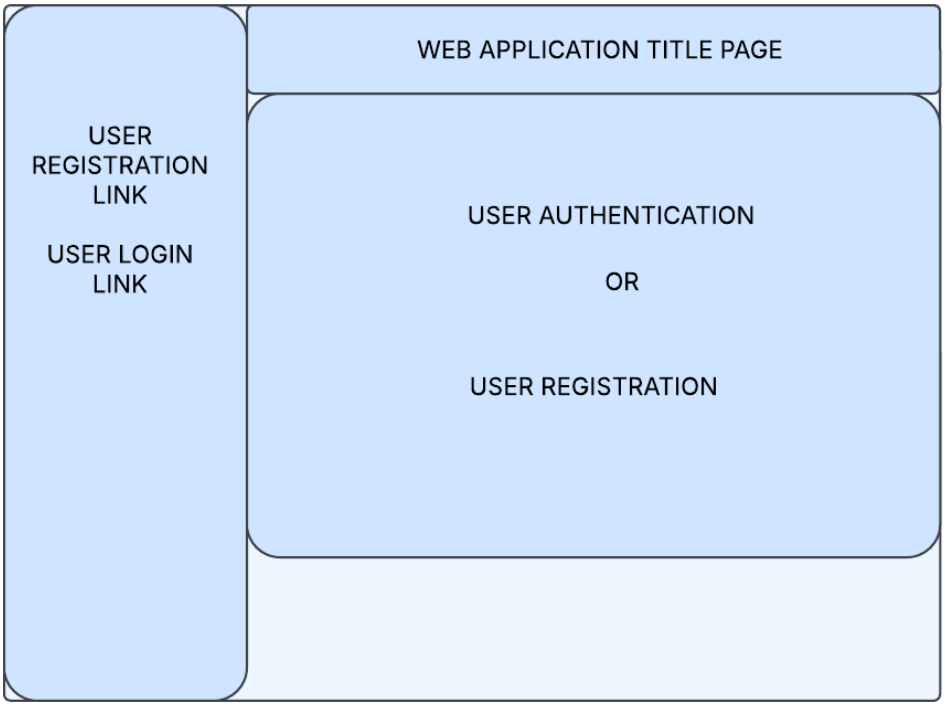


Figure 4.5: User registration wireframe

Chapter 5: System Implementation and Testing

This chapter explores the developments process and technical aspects of the Safety Hazard detection application. We shall look into the architecture, design and implementation aspects of the database. It is the intention of the researcher that this chapter offer information systematic approach taken to develop and implement the safety hazard Detector application. Furthermore, testing methods to evaluate application usability shall be discussed.

5.1 System Implementation

5.1.1 Database

In our research, Structured Query Language (Light Version) (SQLite) was adopted to create and manage our database. It was chosen owing to the scale of application, ease of implementation, reliability and robustness. SQLite is also open-source thereby enhancing accessibility. The database was structured using the SQLite desktop application. The application provided an interface for creation of the data model and configuration of the relationships between the tables. The tables included the detection table, location and user data table as shown in Figure 4.2. The database was normalized in order to conform the Atomicity Consistency Integrity Durability (ACID) properties. This enabled data integrity, redundancy and optimal query performance. Primary and Foreign keys were configured within the SQLite Application to aid in achieving of the ACID properties. The database was directly populated by the application where scripting of the Stream-lit Application was done to ensure this.

5.1.2 Web Portal

The web portal was implemented using Stream-lit Library. Stream-lit is a high-level Python web framework. It was adopted for both front end and backend infrastructure. This library offered reliability and speed in building of a prototype for testing. Stream-lit provides a robust framework for database implementation, session states,and UI rendering. The structure of the web application was also scripted in python file using the Stream-lit library.

(a) Homepage

The Application Main Page, illustrated in Figure 5.1 was designed to provide a simple form that takes in images. The image, once uploaded, is fed to the safety hazard detector. The detector outputs an image with bounding box(s) showing the class of the objects in the image. The image upload form is on the left navigation panel as shown below. Each detection registers the location of the input, image used, time and date of the detection. The data is then stored in the database linked to the user table.

Figure 5.1: Web Application Homepage

(b) User Registration

A user registration form, Figure 5.3, was developed to personalize use of the hazard detector. The user would input their details such as first and last Names, password, username, email and phone number. These data would be linked to a user's activity within the application. It was the desire of the researcher, that a user should have record of their activity for future audit when required. Such records could then be integrated into a Building Information Modeling (BIM) system. In areas where incorrect information is provided, error messages shall be relayed to the user. The user data also enable user authentication using the username and password. The password would be encrypted as a security measure. On accurate completion of the user registration form, the data is stored in the database.

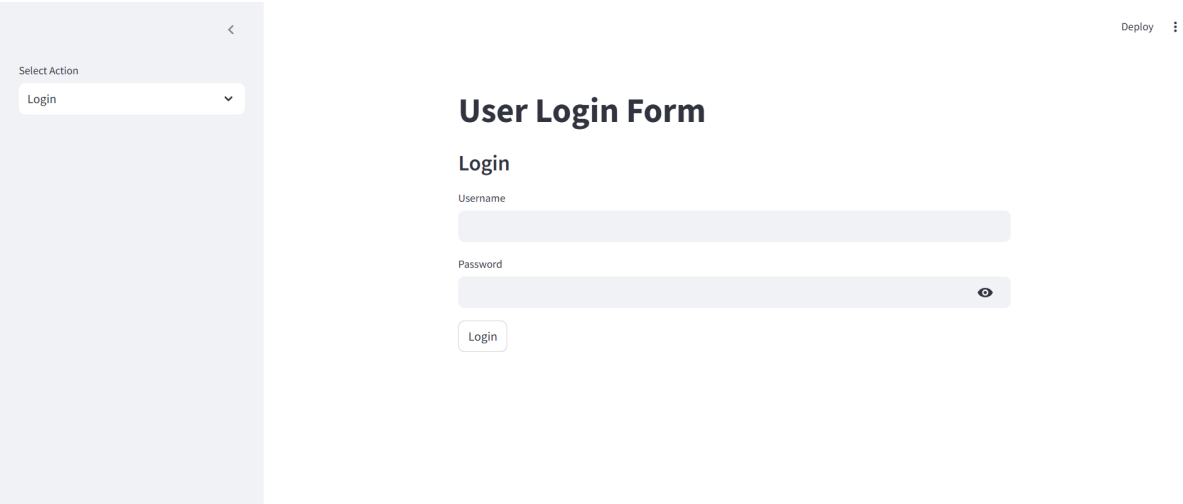


Figure 5.2: Web Application User registration page

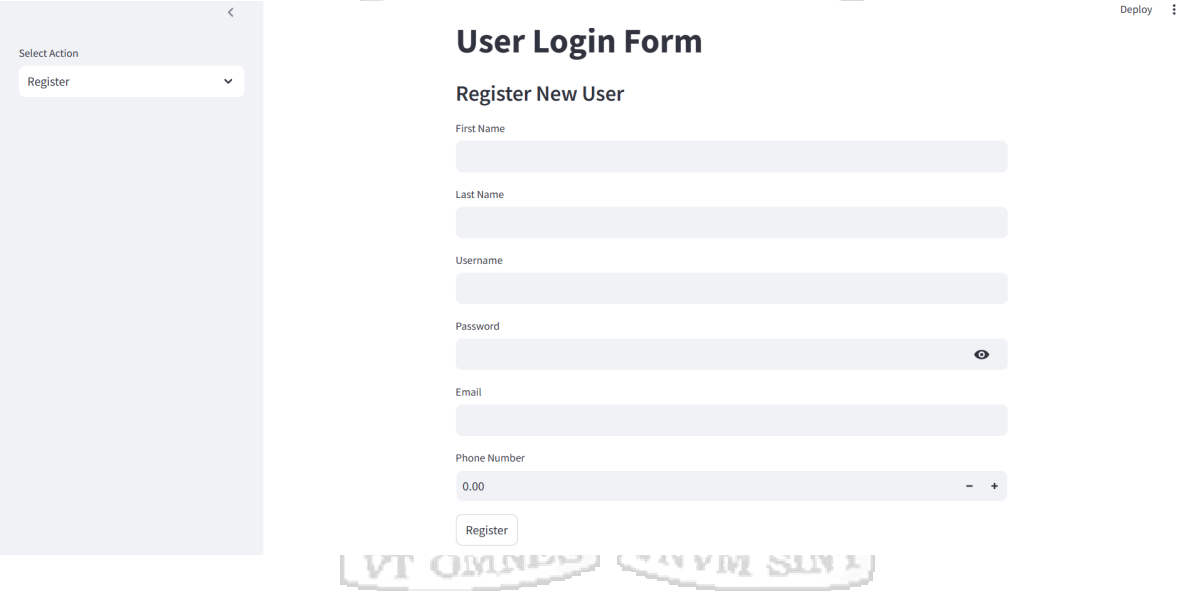


Figure 5.3: Web Application User registration page

5.2 Testing

This stage was implemented to validate the functionality, usability, compatibility, security and efficacy of the safety hazard application. It looked into how our application addressed the problem statement and fulfilled the research objectives. This testing was done to ensure the application would be ready for application in the real world.

5.2.1 Functionality Testing

in this testing the prediction model was tested for using novel images. As shown in Figure 5.1, one image is shown to be tested for object detection. The user registration

and prediction functionality testing was to done to ensure these activities are stored in the database. This storage would be in the correct manner following the expected data type. This testing was done to ensure readiness of the application for real world deployment.

5.2.2 Usability Testing

The usability of the web portal was done by the researcher to ascertain various aspects of the application. This testing included navigating different components of the web application, from user registration, user authentication to uploading and prediction of images for safety hazard detection. The testing outlined potential challenges a user may have on the application and those challenges were addressed during production stage.

5.2.3 Compatibility Testing

Cross functionality across different web browsers was done to assess the versatility and compatibility of the application. The web portal was accessed in different browsers such as Google Chrome, Microsoft Edge, Safari and Firefox. Testers navigated the different components of the application assessed functionality, user-friendliness and performance. Any errors or deviations from expected results was reported back and root-cause analysis done before a solution was formulated.

5.2.4 Security Testing

In our research, robustness of the security protocols was tested to ensure data protection was upheld. The password, which was the main security stronghold was tested for hashing, ensuring its adequacy. The database architecture was also studied to verify secure storage of the data. Additionally, testing was also done to ensure secure API endpoints against security threats.

5.2.5 Validation Testing

Validation testing was done to review the impact of the web application on domain users. This testing was done on the system as a whole to assess whether it could be implemented in the real world. Feedback was provided on how to improve the application for scaled use, such adoption of cloud computing tools. Such feedback would have to implemented outside this research if the application functionality is upgraded.

Chapter 6: Discussion of Results

This chapter provides a walkthrough of the outcomes attained through the stages of the CRISP-DM framework. It includes data understanding, data preparation, modeling, model evaluation and deployment. Useful information from collecting, pre-processing and modeling the data shall be showcased. In this chapter we shall also explore the performance metrics and efficacy of the models in fulfilling the research objectives outlined in Chapter 1, section 1.4

6.1 Data Understanding

The data in our research consisted of 2D images. The images were curated such that each class was well represented in order to balance the dataset. The classes in question were the following:

(a) Fire Hazards

(b) Inhalation Hazards

(c) Electrical Hazards

These classes were chosen based on the ease of visual detection. The assumption of the research is that the model would be connected to a surveillance camera feed installed in a construction site. Such camera equipment tend to be limited in their field of vision as well as resolution, thus may only detect easily recognizable entities the further away they are from the camera. Furthermore, the simplicity of the hazards was a contributing factor to their consideration for this research.

These hazards are common in construction sites, however they do not envisage the multitude of hazards that are encountered. An example is work at height where the hazard is dependent on the juxtaposition of a worker relative to an open ledge or the proximity of a person to moving heavy machinery. This form of computer vision was not included in this study hence such hazards were not considered.

CLASS NAME	COUNT
Electrical-Hazard	362
Fire-Hazard	447
Inhalation-Hazard	466

Figure 6.1: Class Balance of the dataset used for model training from Roboflow tool

The classes, as shown in Figure 6.5, were well balanced with 465, 447 and 361 labels used for inhalation, fire and electrical hazards. This balanced was maintained throughout the data collection and curating process in order to prevent bias in label prediction. The fire hazard labeling entailed images of fire and sparks. Fire in construction is an immediate danger that should be mitigated and controlled as soon as possible. Sparks on the other hand are regular occurrence in construction where welding and grinding of metal takes place. They were considered a fire hazard as they are potential precursor to fire in many situations. Inhalation hazards consisted of dust and smoke. It was noted there is an overlap in class definition of smoke as fire and inhalation hazard, smoke was taken as an inhalation hazard due to its direct effect on an individual's lungs. The electrical hazard consisted of various images, from overloaded sockets to exposed wiring and damaged sockets and extensions. These situations warranted electrical hazards due to the risk of electrical shock. It was also noted that there is a the link between electrical hazards and fire hazards where sparks from damaged electrical components could start a fire.

Images were of varied types with regard to image sizes and quality, shown in Figure 6.2, Figure 6.3 and Figure 6.4. However, the image quality was maintained at an average of 0.05 Megapixels (mp) to enable the model train on images that are of low quality. This was done to enhance model performance in areas of low resolution.

Figure 6.2: Average ratio of images

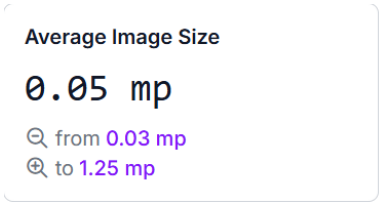


Figure 6.3: Average size of images used in the study

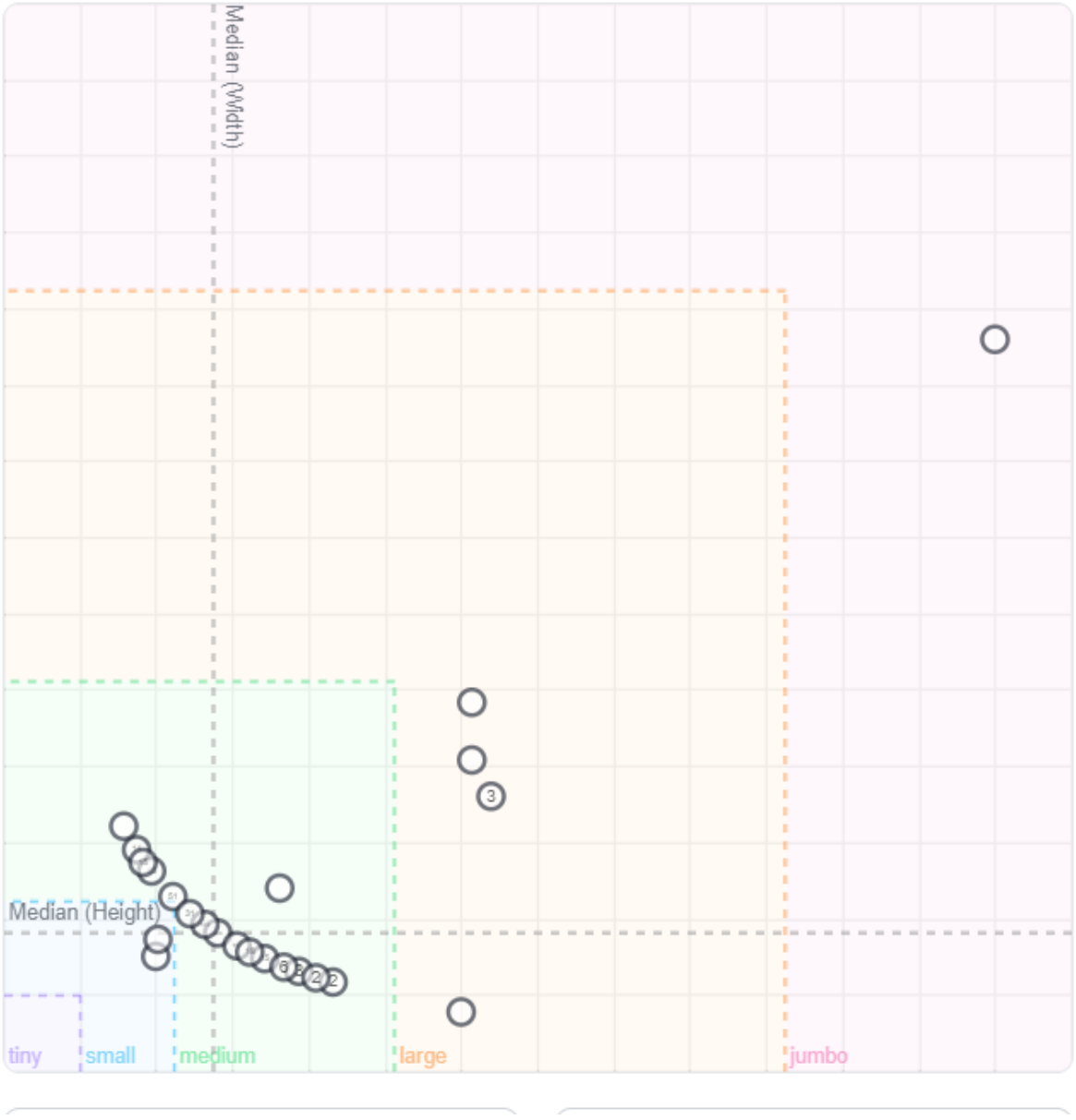


Figure 6.4: Image Size distribution used in the study

6.2 Data Preparation

6.2.1 Image Annotation

Image annotation was carried out using Roboflow tool Dwyer, B., Nelson, J., Hansen, T., et. al. (2024) as mentioned in Chapter 3. The tool proved useful with all images consisting of at least one annotation. Figure 6.5 showcases the annotation distribution.

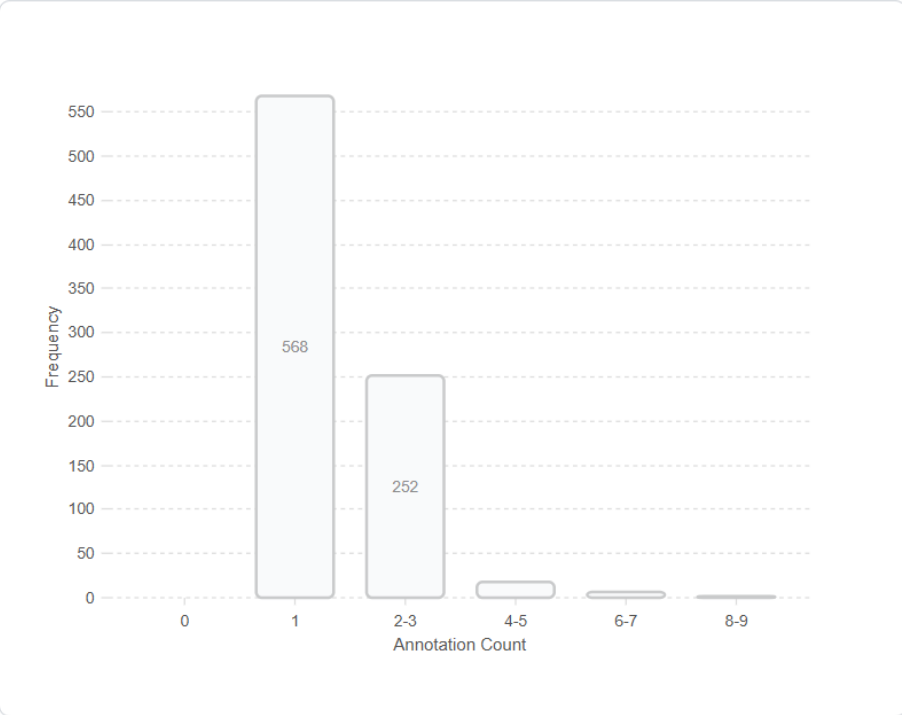


Figure 6.5: Annotation Distribution of the images.

6.2.2 Class Imbalance

The data was curated in a manner that reduces any form of class imbalance. This was done to aid in accurate class discrimination by the model. The three class were balanced in the ratio of 35:35:30 as shown in Figure 6.5:

6.2.3 Image Cleaning

The images were all resized to 640 x 640 mp. This enabled their use in training the YOLO V8 model. For the other two models, Faster R-CNN and SSD, this was not a requirement, as there in-built procedures within the model done to resize the input data to the optimal size. The image sizes were therefore maintained for standardization of image input.

Within the Roboflow tool, the measures employed in image cleaning are summarized in Figure 6.6.

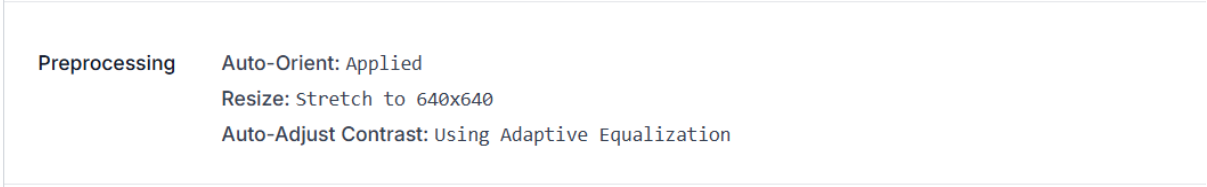


Figure 6.6: Image Preparation.

6.2.4 Image Augmentation

Data augmentation, summarized in Table 6.1, was carried out within the original dataset of 847 images. With the intention of having approximately 1000 images. This was done with the intention of increasing data variation to improve model generalization. The Roboflow tool assisted in this endeavor. Image augmentation was constrained to augmentation of the image without affecting the bounding box location. This was viewed as enough to improve model performance.

Data Augmentation	
Augmentation Technique	Augmentation Value
Rotation	Apply to 15% of images
Hue	Between -15° and +15°
Saturation	Between -25% and +25%
Exposure	Between -11% and +11%
Blur	Up to 2.5px
Noise	Up to 0.1% of pixels

Table 6.1: Summary of Data Augmentation done before model training

6.3 Machine Learning Modeling

This section outlines and discusses the results of the computer vision models adopted for the study.

6.3.1 YOLO Version 8

To analyze the performance of YOLO version 8, a multitude of results had to be considered. In Figure 6.7, the correlation between the ground truth labels and predicted labels is shown. The false negative error of the class labels are 0.84, 0.59 and 0.69 for

the electrical, fire and inhalation hazards. The disparity between the classes is caused by the variability of hazard objects. The electrical hazards are of different types, from exposed wiring to overloaded sockets to damaged extension units. This adds a challenge in classification as the model has to be exposed to a wide variety of data. This is opposite to fire or inhalation hazards which have low variability in shape and pattern with regard to their objects.

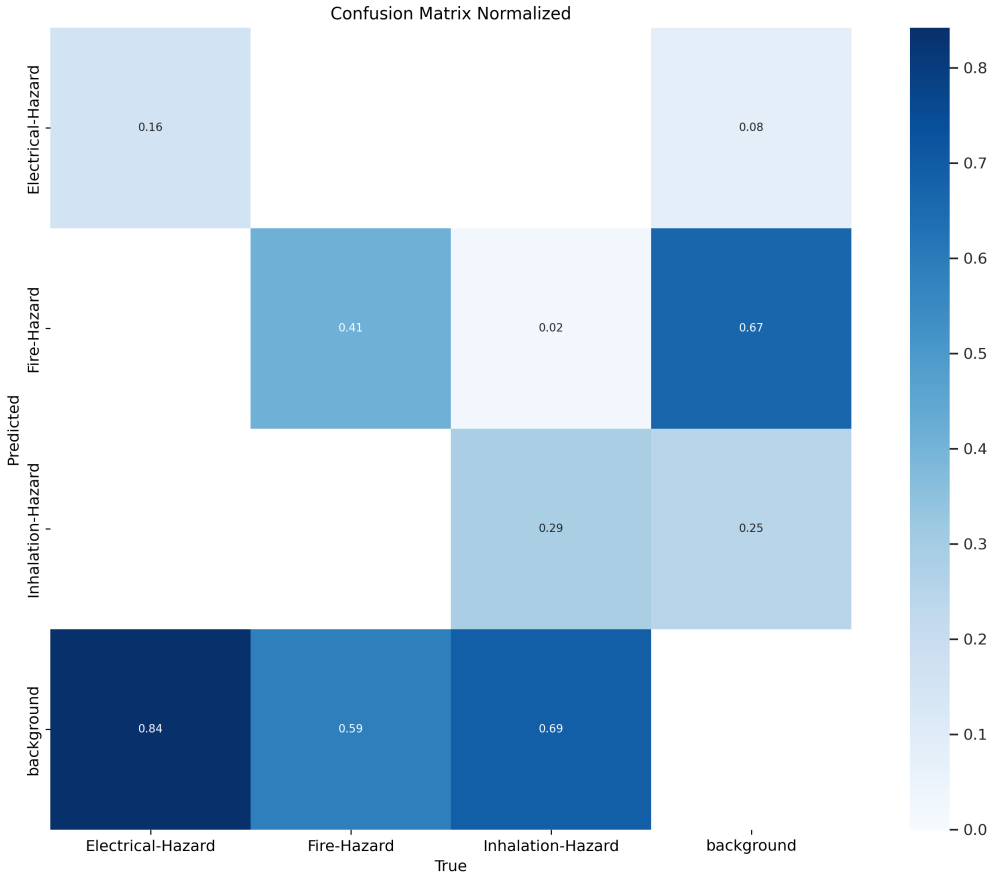


Figure 6.7: Yolo V8 confusion matrix

In Figure 6.8, the curves show a steady increase in precision for all the classes with increase in confidence. However, for fire hazards, there is a sharp decline before a steep increase in precision. This may be due to the presence of noisy data (created through data augmentation) that the model could not adequately analyze. To solve this, additional data could be used to retrain the model for adequate generalization.

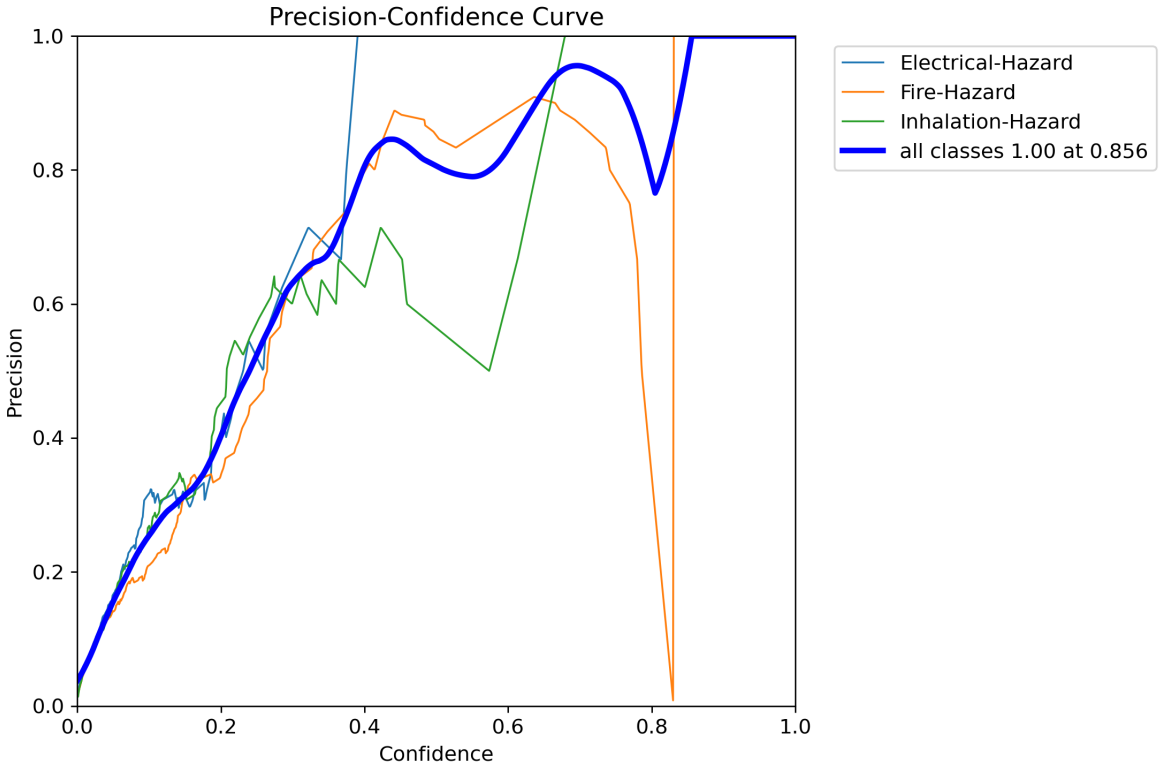


Figure 6.8: Yolo V8 precision-confidence curve

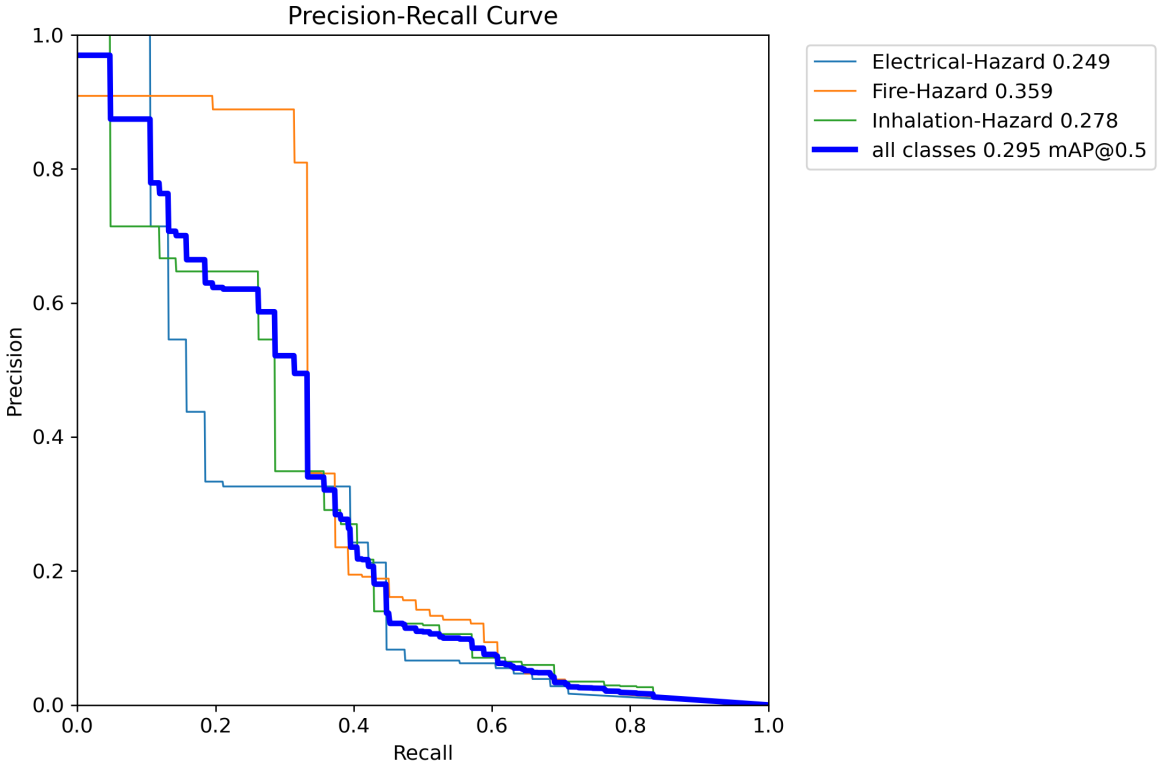


Figure 6.9: Yolo V8 precision-recall curve

In Figure 6.9, the precision recall plot is shown for all classes. The trade off is clearly seen between precision and recall. The electrical hazards show the sharpest decline in the curves. This curve indicates the model ability to distinguish between the classes. The area under the curve was slightly limited however functional for the research objective as multiple predictions are made for the target object. It is assumed the model would have video input hence the even with recall of 30%, this should be enough to detect a safety hazard. Furthermore, the hazards in consideration have little to no mobility hence the probability of prediction is sufficient.

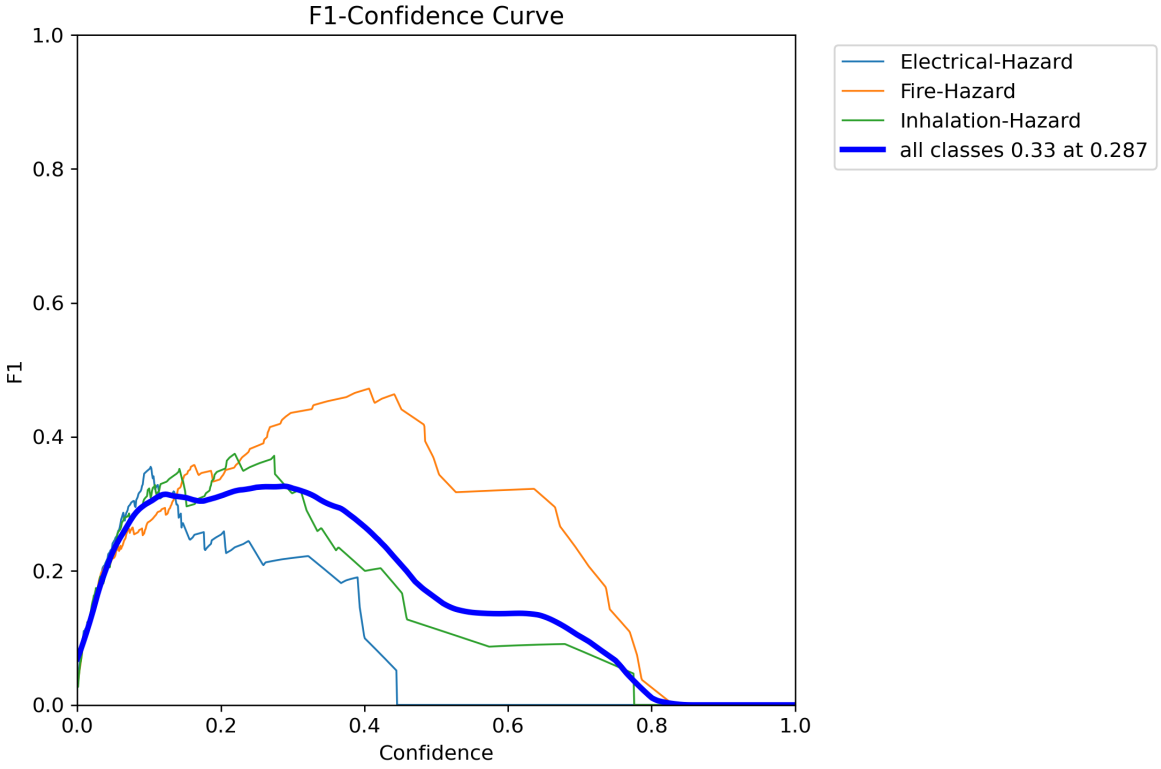


Figure 6.10: Yolo V8 F1 Recall curve

In Figure 6.10, the average F1-score is 33 %. This value takes into account the precision and recall of the model. This was model confidence of 28.7 %. This is acceptable due to the static of the hazards in question as previously mentioned.

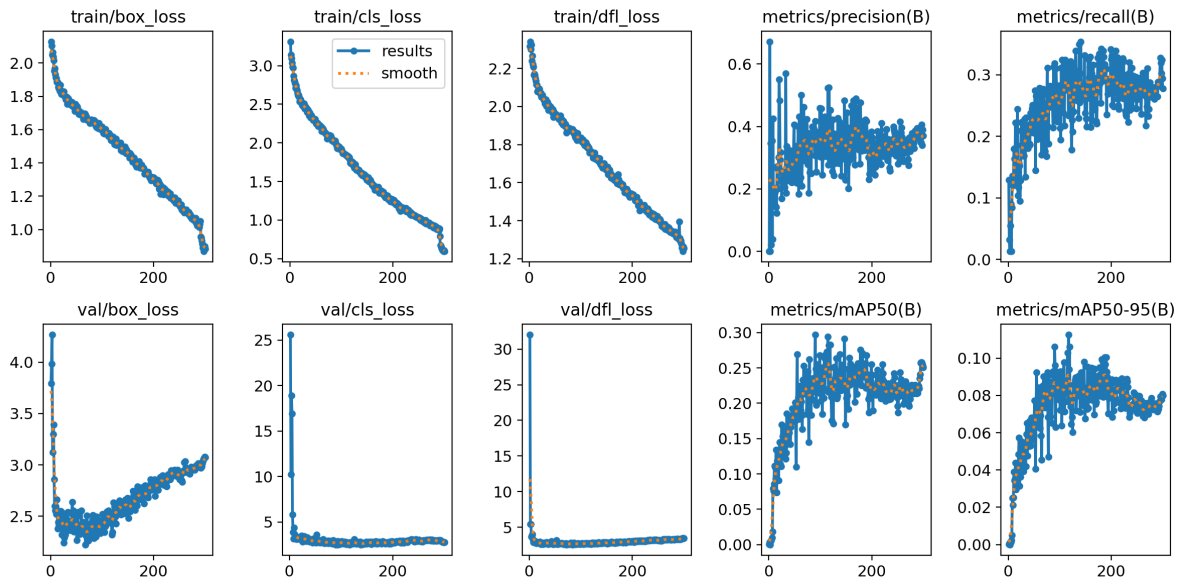


Figure 6.11: Yolo V8 Overall Results

Overall results from training the YOLO version 8 algorithm are shown in Figure 6.11. The numerous graphs show the training and validation loss as well as the mean average precision mAP at IOU of 50% and 95%. The mAP at 50% is 30% while that at 95% is 15%. This is acceptable given the nature of the research. The hazards in question tend to be large in size and easy to detect. The losses show a steep decline after a few epochs, for the validation loss and a gradual decline for the training loss. This elaborates the algorithm's attempt at generalization. Furthermore, the model is pre-trained, hence less epochs are necessary for training the model adequately.

However, the validation box loss began to increase as the number of epochs increased. This was a sign of overfitting and the epoch number was limited. The recall of the model showed a steady rise as the number of epochs increased. However, as the epochs increased in number there was flattening of the curve, showcasing diminishing returns from the model.

6.3.2 Faster R-CNN

In this section we shall discuss the results from the Faster R-CNN model. In Figure 6.12 we see the validation precision obtained during evaluation of the model. As the training progressed, there was a sharp increase in precision to 18% during the first few epochs. This was expected as the model is pre-trained. As such finer details such as edges and curves

are already learned by the model parameters. This accuracy slightly reduces to about 16% after which it stabilizes and remains constant. This signifies that the training epochs were enough to train the model and any increase in accuracy would require additional amounts of data.

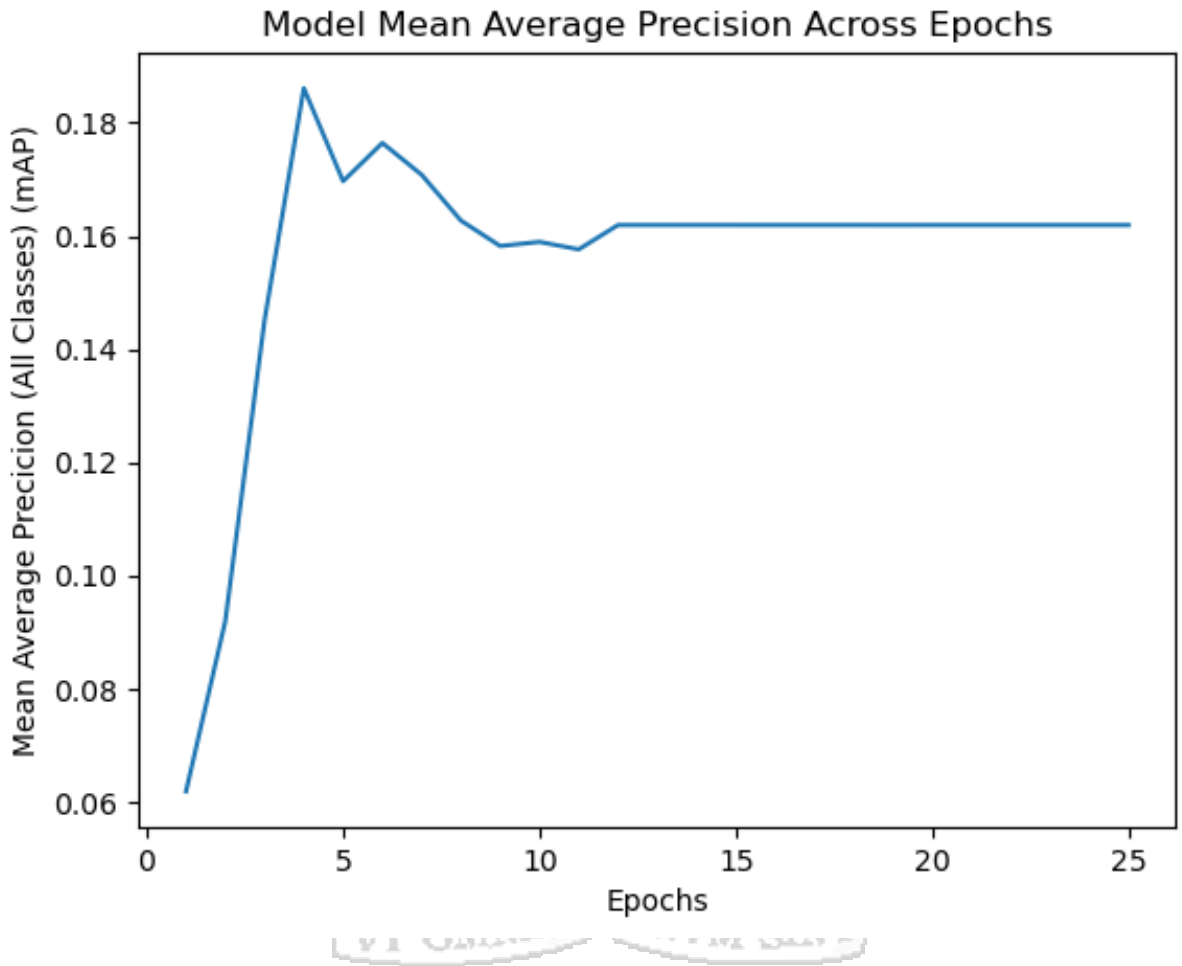


Figure 6.12: Faster R-CNN Mean Average Precision Results mAP

In Figure 6.13, we see higher precision values for detection. This is due to the less stringent IOU requirements by the metrics. This metric is also useful owing to the large sizes of the objects to be detected.

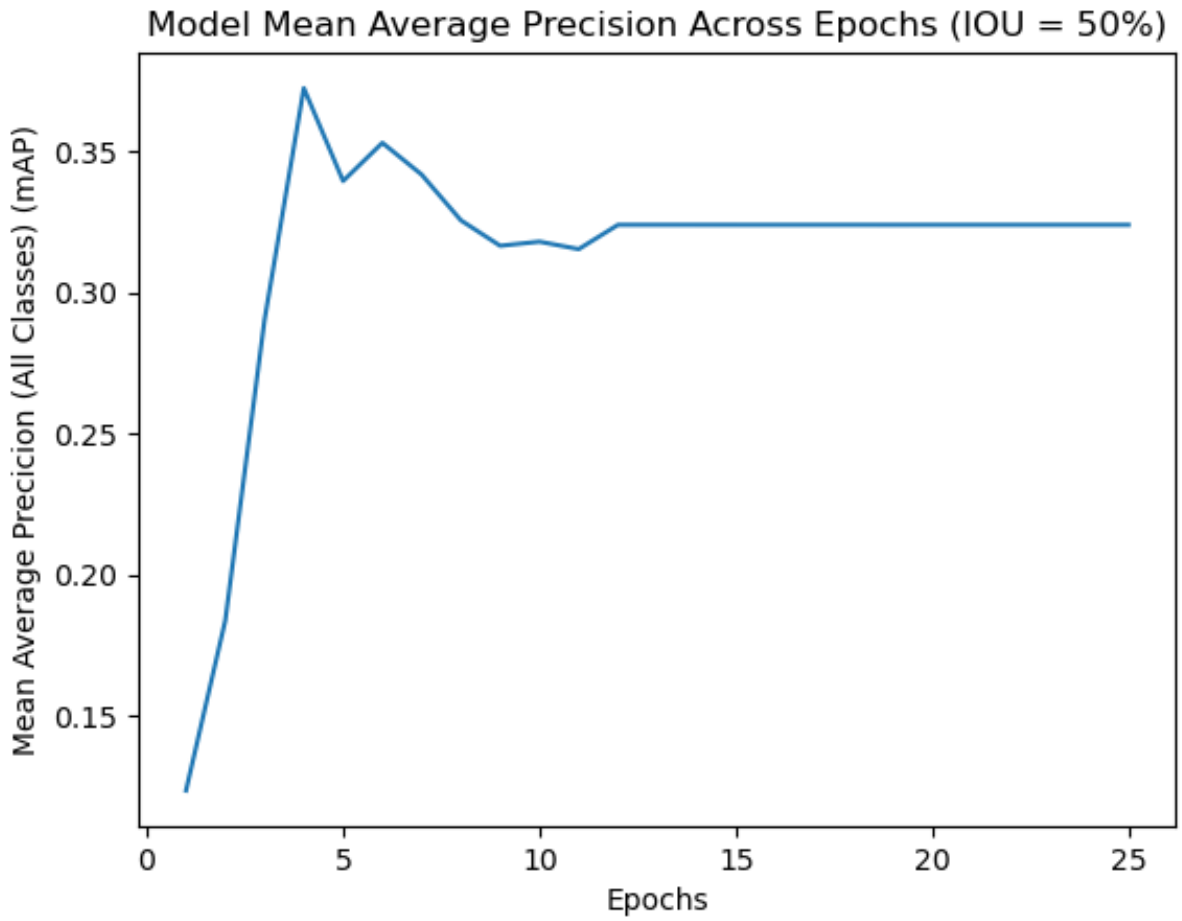


Figure 6.13: Faster R-CNN Mean Average Precision Results mAP at IOU of 50%

All class mAP was measured as done in the YOLO algorithm. It was the intention of the research to know how well the model recognizes the different classes. This would be pivotal in the further training to ensure equality of class distinction. This is summarized in Figure 6.14. It can be seen that fire hazards are better distinguished with mAP stabilizing at 25%. This is due to the low variability of the object. Electrical hazards showed low mAP stabilizing at 10%, this was theorized to be due to high variability of the object shape and size. The inhalation hazards also showed lower mAP at about 13% where such hazards were also more difficult to perceive in certain images owing to the lower contrast differences with the background. A good example is detecting brown dust on a sunny day in a desert area.

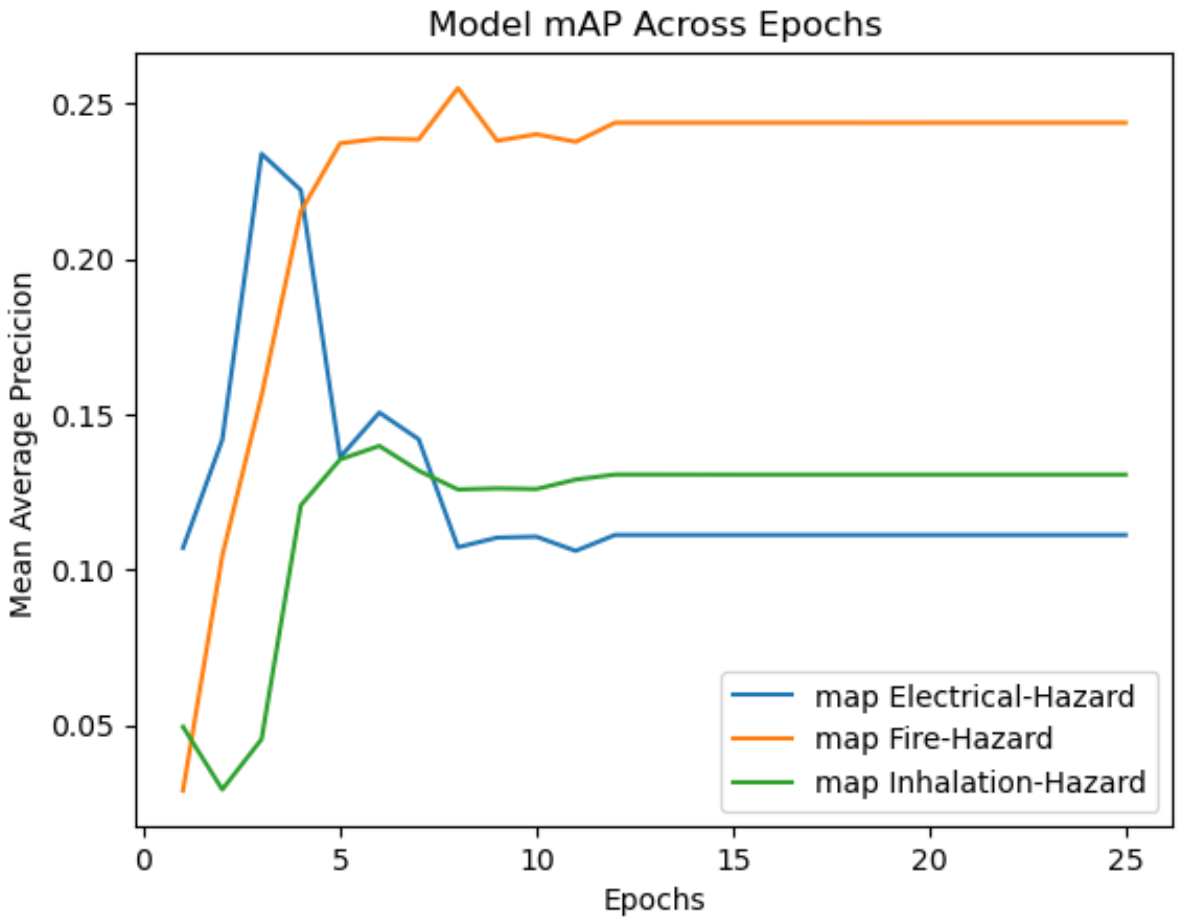


Figure 6.14: Faster R-CNN Mean Average Precision Results mAP for different classes

Recall was another metrics considered in evaluating the Faster R-CNN model. Our study considered the mean average recall for the fist 10 detection as the standard metric. This was chosen taking into account the nature of hazards being detected. The object are static hence some level of recall can be seceded. False negatives can be tolerated with moderate cost to the end user. In Figure 6.15, this is shown where there average recall sharply rises and before stabilizing at 45%.

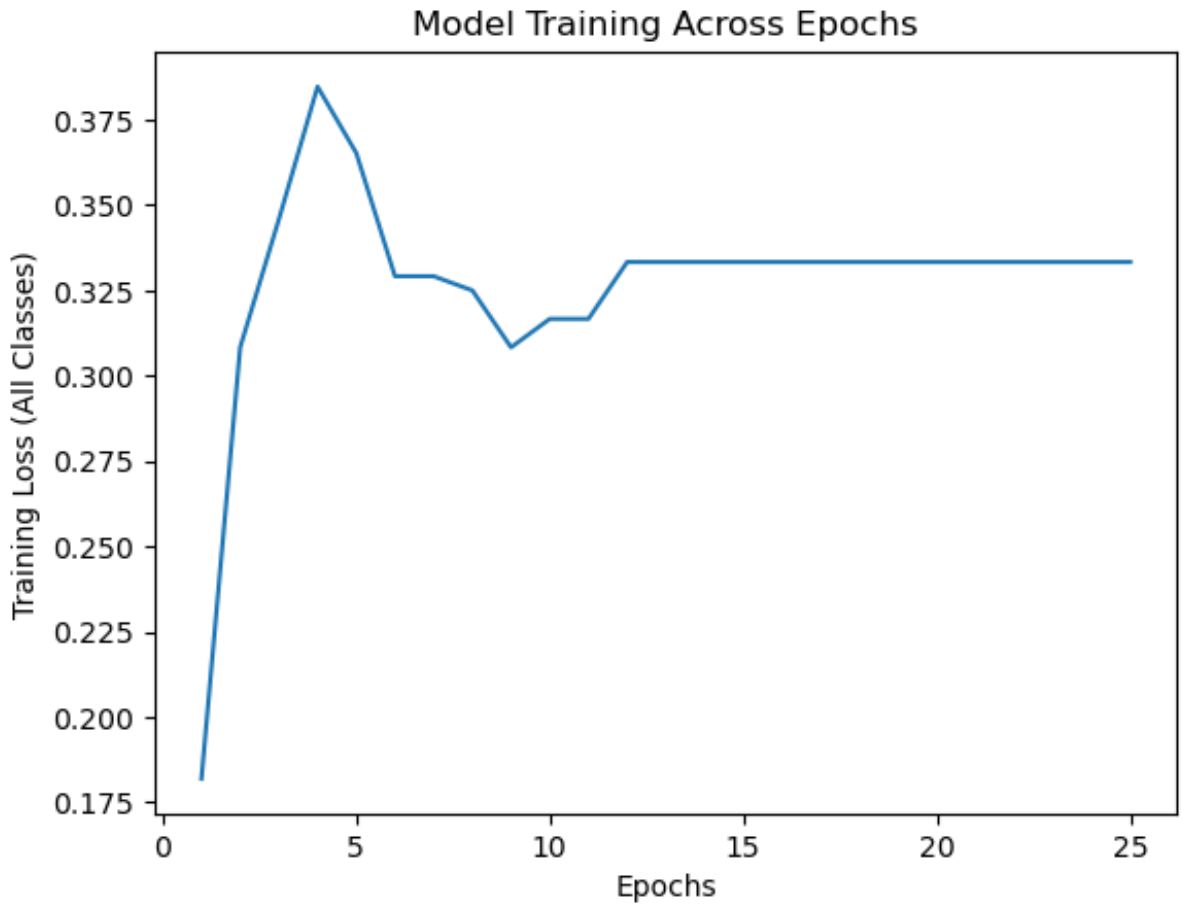


Figure 6.15: Faster R-CNN Mean Average Recall mAR

The final metric that was considered was the F1-score. In Figure 6.16, the value is seen to sharply increase to 38% before stabilizing at 33%. This score indicates the ability of the model to distinguish between the different classes. Owing to the balance of the dataset, this was expected, where the F1-score is adequate, even higher than precision values.

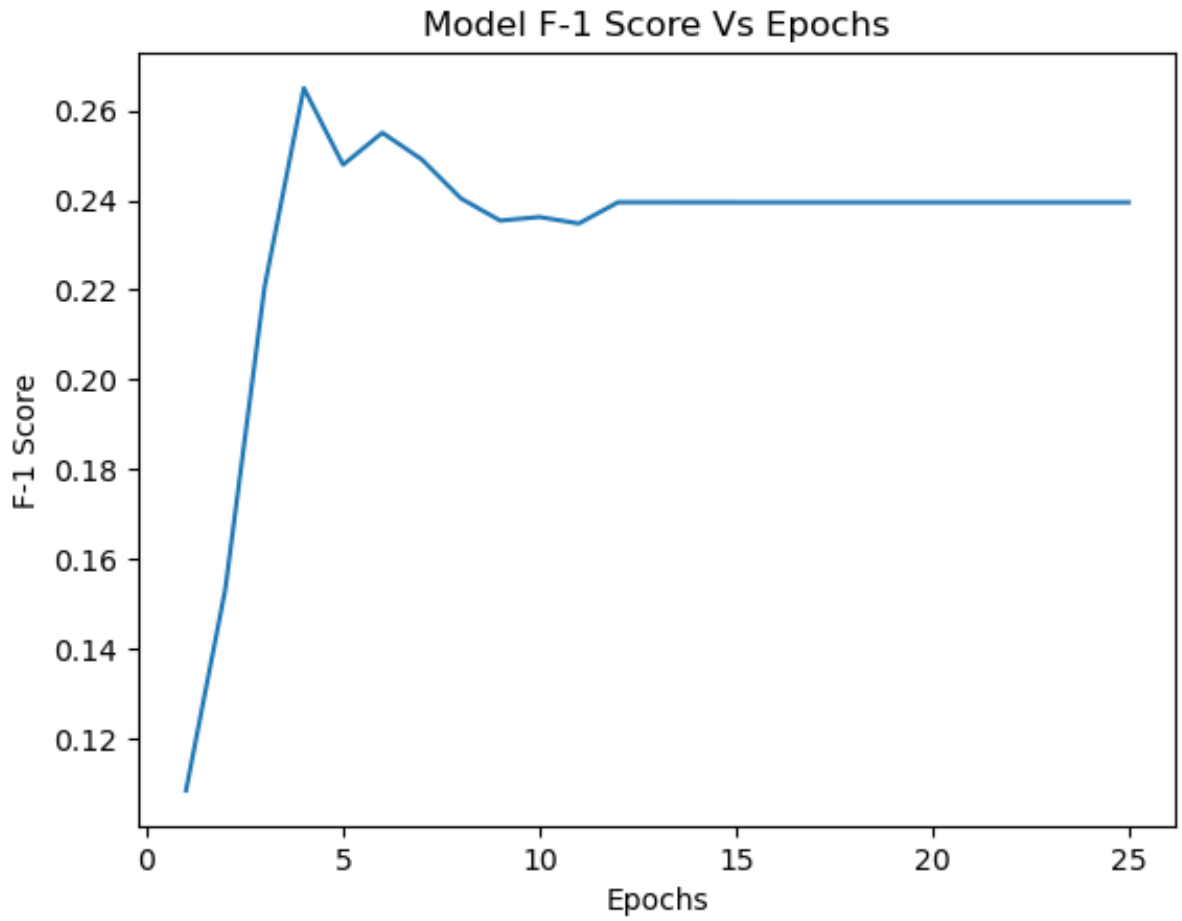


Figure 6.16: Faster R-CNN Mean F1-Score

6.3.3 Single Shot Detector SSD

The last model considered for this study was the Single Shot Detector. The metrics used for this model are similar to those of the Faster R-CNN. This is due to the similarity of libraries used for metrics generation. In Figure 6.17, the mAP is shown to rise exponentially during the first few epochs as expected. It rises to about 22% before stabilizing at 19%. This slightly higher than the performance of the Faster R-CNN model. This can be attributed to the fact that SSD model is configured for low resolution images (Liu et al., 2016). This makes it better attuned for noisy data. The steep rise in accuracy is attributed to the pre-training of the model using the COCO dataset, similar to the R-CNN model.

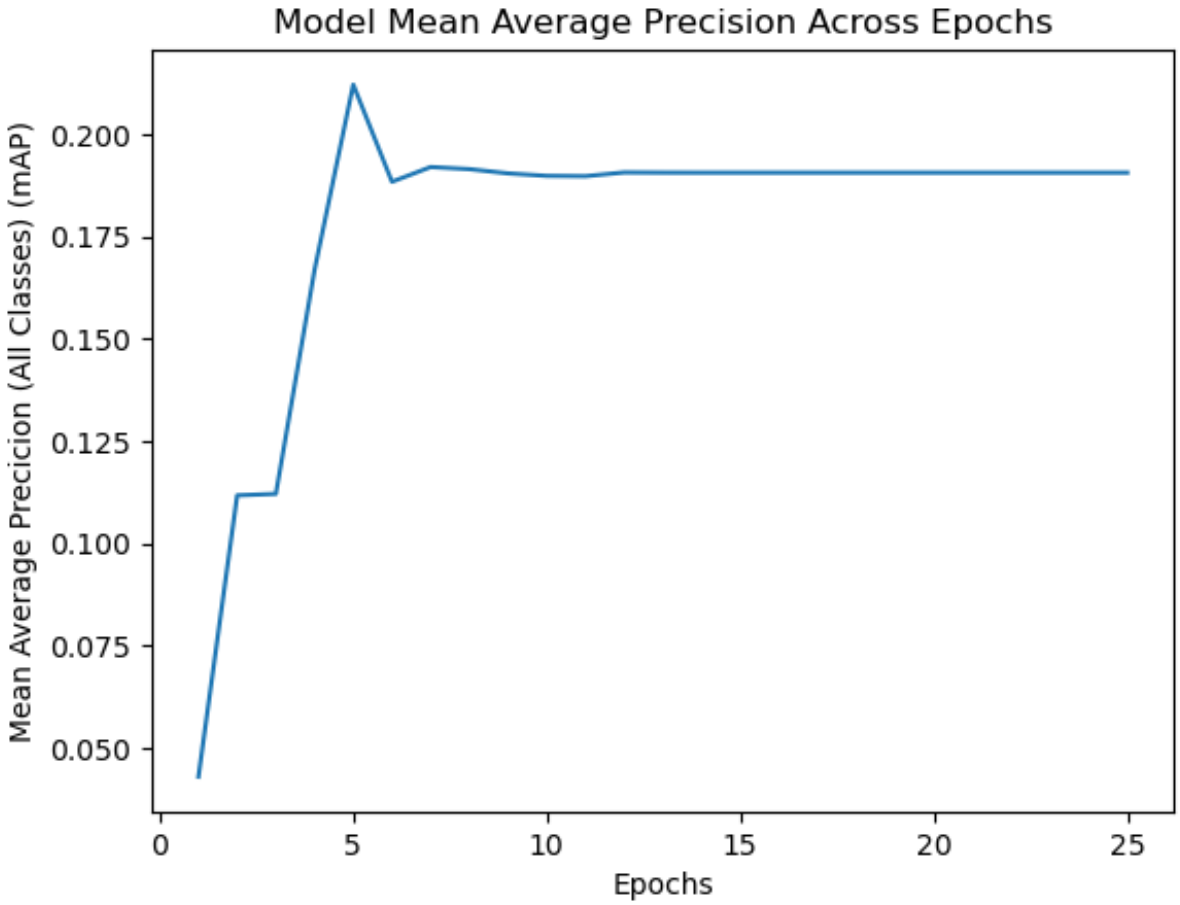


Figure 6.17: SSD Mean Average Precision Across All Classes

The mAP for IOU of 50% as shown in Figure 6.18 was also considered. The trend of the accuracy is similar to that of the mAP for all classes. The mAP steeply rises to approximately 43% and stabilizes at 40%. The reason for this trend is similar to what was seen in the average mAP previously discussed.

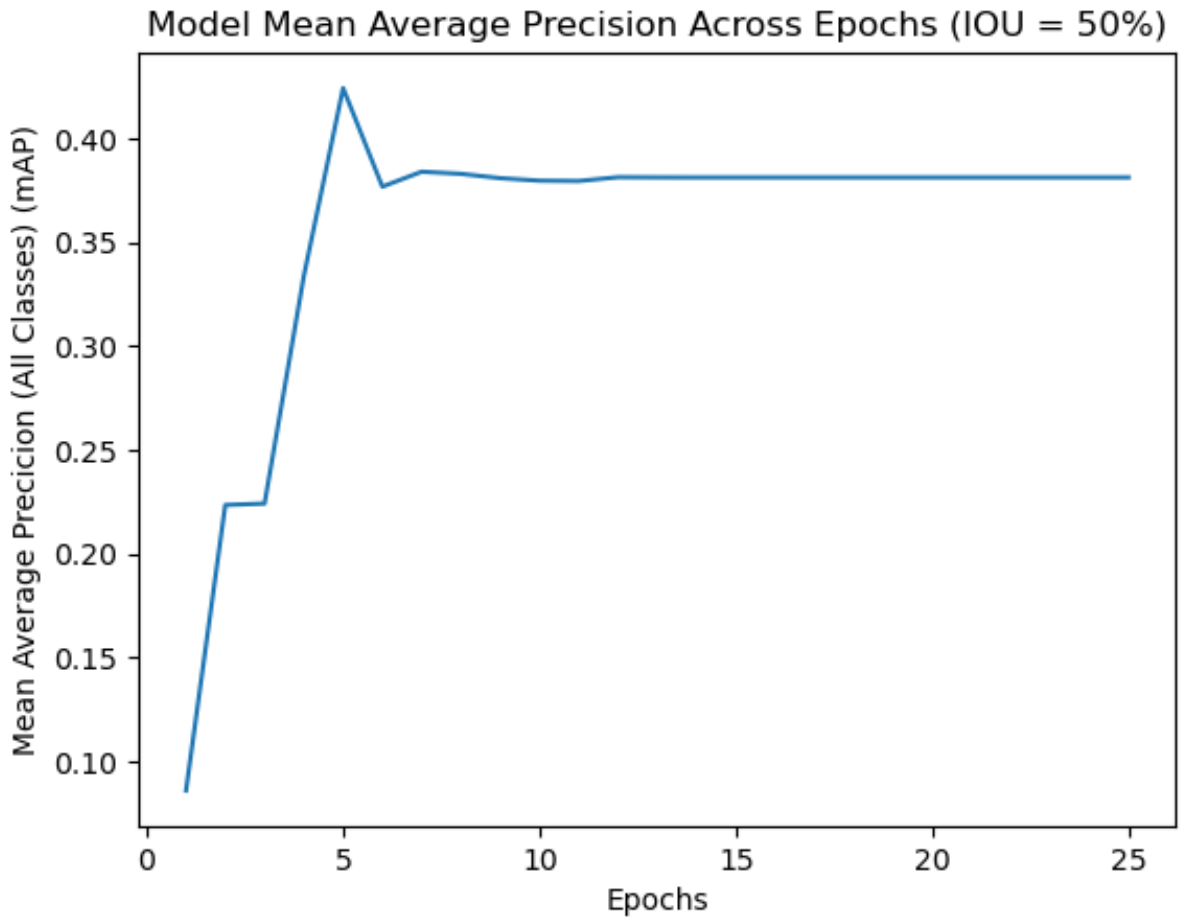


Figure 6.18: SSD Mean Average Precision Across All Classes for IOU of 50%

Figure 6.19 shows the performance of each class. The fire hazards are easily detected as seen in the previous models with mAP reaching over 25%, this is followed by electrical hazards reaching precision levels of 20%. Lastly, the inhalation hazard follow with mAP of 12%. The fire hazards have better precision owing to the differential contrast seen in fire objects. Inhalation hazards object are opposite to this, where for proper distinction, more training may be required.

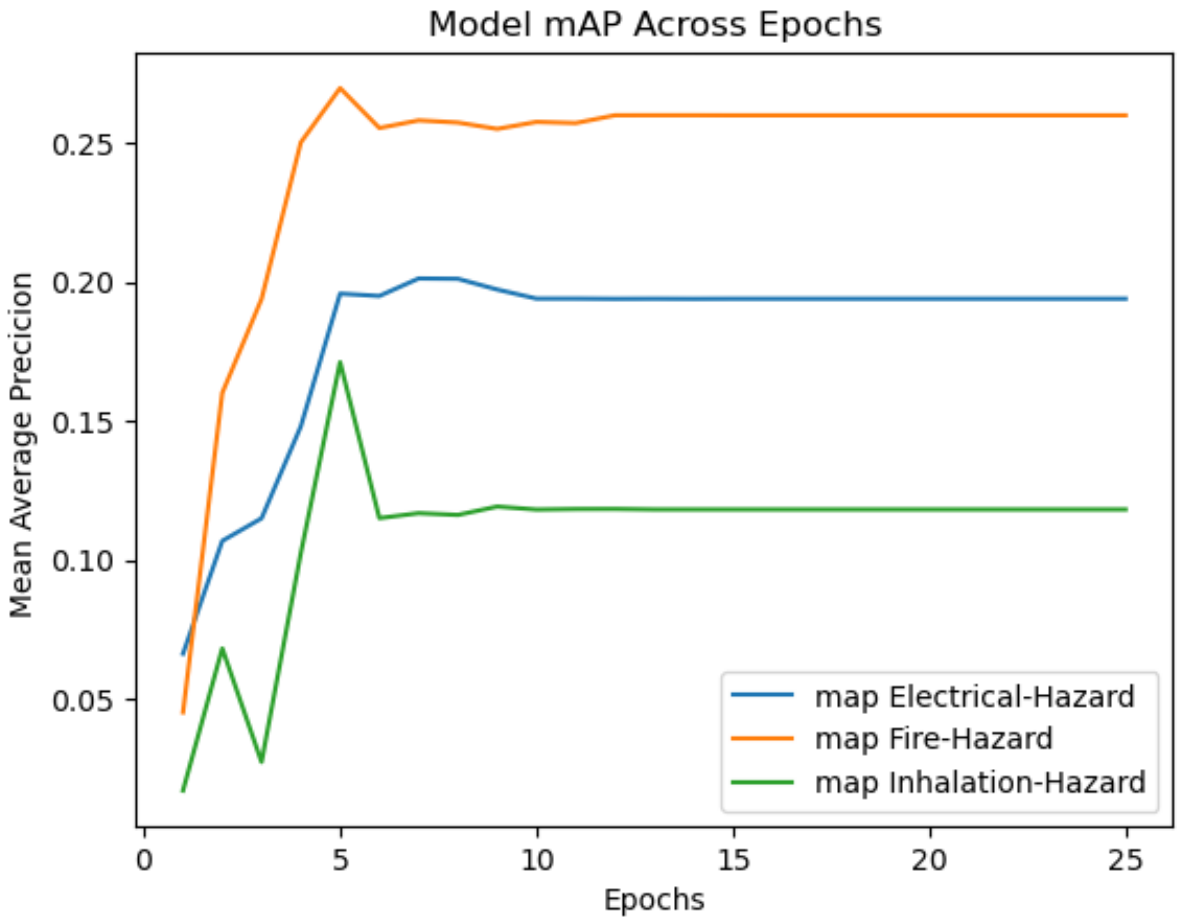


Figure 6.19: SSD Mean Average Precision for each of the Classes

The SSD model mean average recall was also investigated in Figure 6.20. Its performance was similar to the Faster R-CNN model. The mean recall showed a steep increase to 38%, before stabilizing at 35%. Similar to all other metrics, the curves were similar in trend and trajectory. Any further training in the model led to no increase in recall, hence the number of epochs was curtailed to 25.

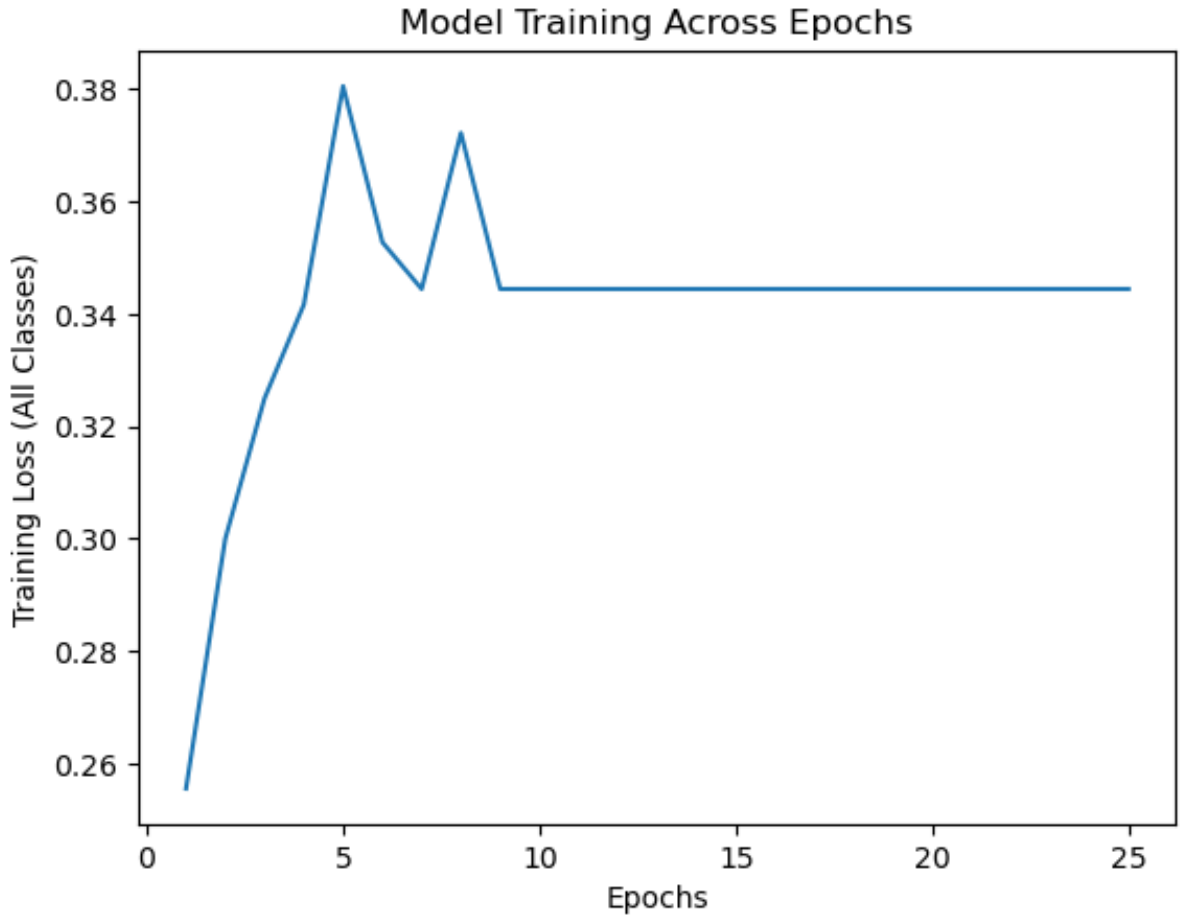


Figure 6.20: SSD Mean Average Recall (averaged across all classes)

The final metric that was utilized for the SSD model evaluation was the F1-score, as done in the previous models. In Figure 6.21, The F1-score rises sharply to 27% before stabilizing at just above 25%. The mAR was taken for 100 detections.

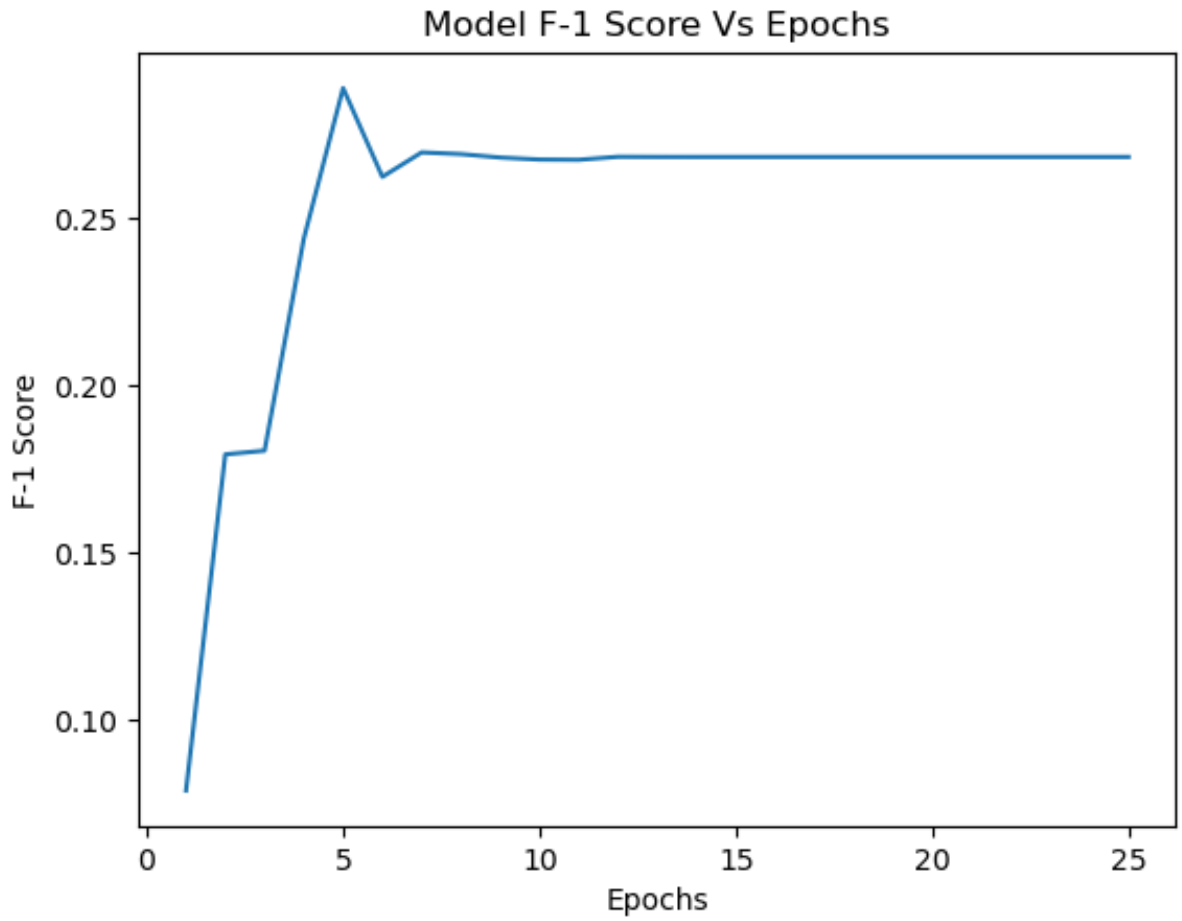


Figure 6.21: SSD F1-Score (averaged across all classes)

6.3.4 Model Optimization

Model optimization was done independently for all models. All models showed exponential reduction in training loss and increase in validation precision. For the YOLO version 8 model, this can be seen in Figure 6.11. For the rest of the models the training loss are shown in Figure 6.22 and Figure 6.23.

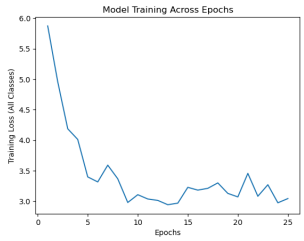


Figure 6.22: Training Loss -Faster R-CNN

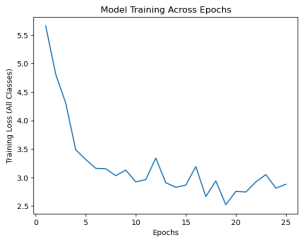


Figure 6.23: Training Loss- SSD

To attain optimization shown above, the following parameters in Table 6.2 and Table 6.3 were utilized.

Model Hyper-parameters	
Hyper parameter	Value
Image size	640x640
Colour Channels	3
Batch size	32
Number of Epochs	300
Optimizer	Adam
Learning Rate	0.01

Table 6.2: Summary of the hyper-parameters utilised for YOLO v8 model training

Model Hyper-parameters	
Hyper parameter	Value
Image size	640x640 (300x300 for SSD)
Colour Channels	3
Batch size	32
Number of Epochs	25
Optimizer	Adam
Learning Rate	0.0005
Learning Rate Decay	0.000001

Table 6.3: Summary of the hyper-parameters utilised for Faster R-CNN and SSD model training

For the Faster R-CNN and SSD, a learning scheduler was adopted taking in to account the weight decay to prevent the model from settling in local minima during backpropagation.

6.4 Summary

The results indicate that YOLO version outperformed in consideration of the mean average precision corroborating the work of Reis et al. (2024) on the applicability of the model in transfer learning. This metric bears more weight as it is the standard for performance of the object detection models.

Chapter 7: Conclusions, Recommendations and Future Work

7.1 Conclusion

This research highlights on the role computer vision can play in the construction industry. It addresses a pain point in the industry, safety. Computer vision can be leveraged to detect hazards in construction sites, aiding safety teams that carry out this role. Through the use of computer vision models, safety hazard detection can be automated and therefore scaled to many sites. This solution also enables the safety teams focus on hazard mitigation, reducing the need for numerous site walkabouts needed to detect hazards. Furthermore, other stakeholders such as construction workers, engineers, architects, project managers can be protected by this technology.

The application of this model to a web application as shown elucidates the potential for scalability and adoption in other set-ups besides construction. An image or video feed can be uploaded to a simple web application and the model can begin providing value to the stakeholder. It is the belief of the researcher that adoption of computer vision technology in the realm of safety may not only save lives but preserve the value for stakeholders and players in the construction industry. This can then shall usher an era of safe construction sites with minimal safety incidents spearheaded by technology as shown in this research.

7.2 Recommendations

The study recommends the adoption of the study in realm of construction as guided below. Stakeholder engagement will need to be carried out, where the owner of the construction project is advised on the utility of the model. This is followed by fostering of collaboration with the construction consultants, especially the safety professional. This is crucial for success of the solution. Finally, the contractor will need to be brought onboard as he shall provide the equipment necessary for the model to work; surveillance cameras or smart phones to create input for the model to infer. It is important for the model to be integrated to a web or mobile application.The camera (located onsite) inputs data directly to the model and the output appears in the application. This should enable visibility by all stakeholders regardless of location. Such an application may be embedded in a BIM system for consolidation of the model output which can then also be recorded as historical information.

Finally, it is crucial for advocacy of the solution to industry players and the public in general. Framework policies and advocacy by institutions on the need for leveraging technology to tackle pain-points in construction such as safety may prove valuable.

7.3 Future Works

There are numerous avenues for extension of this study. The number of safety hazards adopted for the study is limited. Therefore, there are many hazards such as fall from height, worker-machine collision, injury from falling objects and many more that can be adopted for research. Simple object detection, with bounding boxes, was done for this study, however, the exploration of advanced computer vision techniques such as object segmentation, key-point estimation may provide returns. The aforementioned additional hazards may be detected using these computer vision techniques.

Finally, there is the potential for image-text transformer models that could be used to generate safety report based on hazard detection. Pre-trained computer vision and language models can be combined and fine-tuned to this task. This may provide significant value addition to stakeholders in construction endeavors. These directions may prove useful in the expansion of technology adoption in the field of safety and to a larger extend, construction.

Bibliography

(2001). *Developing Highly Effective Construction Safety Professionals*, volume All Days of *Professional Development Conference and Exposition*.

Diwan, Parag Kothari, D. (2015). Role of automation and robotics in semiconductor industry. *IETE Technical Review*, 7:368–377.

Dwyer, B., Nelson, J., Hansen, T., et. al. (2024). Roboflow (version 1.0) [software].

Fang, W., Love, P. E., Luo, H., and Ding, L. (2020). Computer vision for behaviour-based safety in construction: A review and future directions. *Advanced Engineering Informatics*, 43:100980.

Fang, W., Zhong, B., Zhao, N., Love, P. E., Luo, H., Xue, J., and Xu, S. (2019). A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network. *Advanced Engineering Informatics*, 39:170–177.

Ghosh, A. (2024). Yolo version 8 architecture.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587.

Guo, B. H., Zou, Y., Fang, Y., Goh, Y. M., and Zou, P. X. (2021). Computer vision technologies for safety science and management in construction: A critical review and future research directions. *Safety Science*, 135:105130.

ILO (2005). A global alliance against forced labour: Global report.

Kim, H., Kim, K., and Kim, H. (2016). Vision-based object-centric safety assessment using fuzzy inference: Monitoring struck-by accidents with moving objects. *Journal of Computing in Civil Engineering*, 30(4):04015075.

Koc, H., Erdoğan, A., Barjakly, Y., and Peker, S. (2021). Uml diagrams in software engineering research: A systematic literature review. *Proceedings*, 74:13.

Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*, 25.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). *SSD: Single Shot MultiBox Detector*, page 21–37. Springer International Publishing.

Luo, H., Wang, M., Wong, P. K.-Y., and Cheng, J. C. (2020). Full body pose estimation of construction equipment using computer vision and deep learning techniques. *Automation in Construction*, 110:103016.

Muñoz-La Rivera, F., Mora-Serrano, J., and Oñate, E. (2021). Factors influencing safety on construction projects (fscps): Types and categories. *International Journal of Environmental Research and Public Health*, 18(20).

Nguyen, V. N., Jenssen, R., and Roverso, D. (2018). Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning. *International Journal of Electrical Power Energy Systems*, 99:107–120.

Nyerere, R. K. . J. (2016). Occupational accident patterns and prevention measures in construction sites in nairobi county kenya. *American Journal of Civil Engineering*, 4(5):254–263.

Padilla, R., Netto, S. L., and da Silva, E. A. B. (2020). A survey on performance metrics for object-detection algorithms. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 237–242.

Reis, D., Kupec, J., Hong, J., and Daoudi, A. (2024). Real-time flying object detection with yolov8.

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149.

Seo, J., Han, S., Lee, S., and Kim, H. (2015). Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics*, 29(2):239–251. Infrastructure Computer Vision.

Suman Paneru, I. J. (2021). Computer vision applications in construction: Current state, opportunities challenges. *Automation in Construction*, 132:103940.

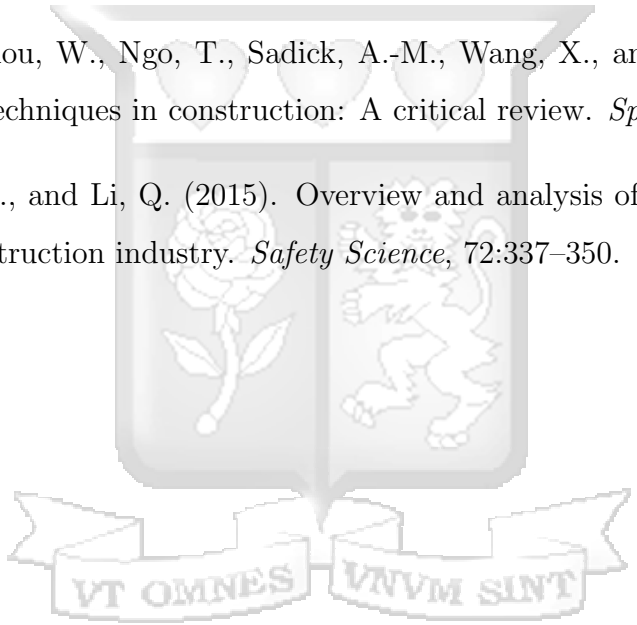
Sun, L. (2016). Learning object model via segment-layout topic. In *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, pages 590–595.

Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.*, 2018:7068349.

Xu, S., Wang, J., Shou, W., Ngo, T., Sadick, A.-M., and Wang, X. (2021). Computer vision techniques in construction: A critical review. *Arch. Comput. Methods Eng.*, 28(5):3383–3397.

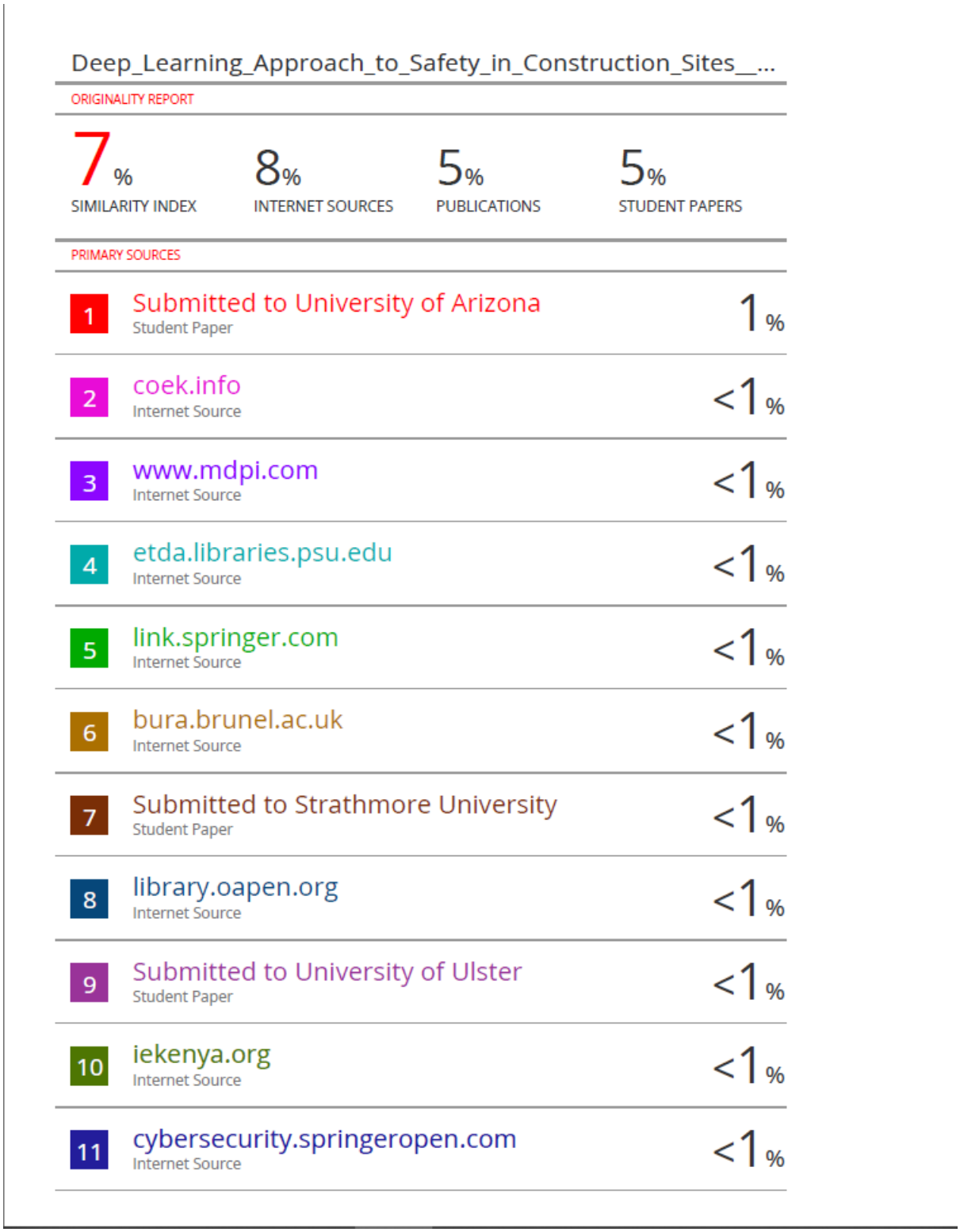
Xu, S., Wang, J., Shou, W., Ngo, T., Sadick, A.-M., Wang, X., and Wang, X. (2020). Computer vision techniques in construction: A critical review. *Springer Netherlands*.

Zhou, Z., Goh, Y. M., and Li, Q. (2015). Overview and analysis of safety management studies in the construction industry. *Safety Science*, 72:337–350.



Appendices

Appendix A: Similarity Report



Deep_Learning_Approach_to_Safety_in_Construction_Sites__...

ORIGINALITY REPORT

7%	8%	5%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Arizona Student Paper	1%
2	coek.info Internet Source	<1%
3	www.mdpi.com Internet Source	<1%
4	etda.libraries.psu.edu Internet Source	<1%
5	link.springer.com Internet Source	<1%
6	bura.brunel.ac.uk Internet Source	<1%
7	Submitted to Strathmore University Student Paper	<1%
8	library.oapen.org Internet Source	<1%
9	Submitted to University of Ulster Student Paper	<1%
10	iekenya.org Internet Source	<1%
11	cybersecurity.springeropen.com Internet Source	<1%

12	mspace.lib.umanitoba.ca Internet Source	<1 %
13	aranne5.bgu.ac.il Internet Source	<1 %
14	5cad8ab8-943d-4294-b849-5f790e377072.filesusr.com Internet Source	<1 %
15	dokumen.pub Internet Source	<1 %
16	iopscience.iop.org Internet Source	<1 %
17	Belchior, Lúcia Madeira. "Churn Prediction in Online Newspaper Subscriptions", Universidade NOVA de Lisboa (Portugal), 2024 Publication	<1 %
18	Submitted to University of Western Sydney Student Paper	<1 %
19	su-plus.strathmore.edu Internet Source	<1 %
20	"Advances in Information Technology in Civil and Building Engineering", Springer Science and Business Media LLC, 2024 Publication	<1 %
21	Submitted to Segi University College Student Paper	<1 %

Exclude quotes Off Exclude matches < 25 words

Appendix B: Ethical Clearance Confirmation

Strathmore UNIVERSITY

20th February 2025

Mr Maina Anthony,
anthony.maina@strathmore.edu

Dear Mr Maina,

RE: Deep Learning Approach to Safety in Construction Sites

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2554/25**. The approval period is from **20th February 2025 to 19th February 2026.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu