Ensuring Safety for Kindergarten Students: A Real-Time IoT-Based Alarm and Monitoring System at Manalupang San Vicente Elementary School

A Design Project
Presented to the Faculty
of the Computer Engineering Department
College of Engineering
BATANGAS STATE UNIVERSITY
Alangilan Campus, Batangas City

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Science in Computer Engineering

DAWIS, LOURYDEM B.

JARO, JAN MILLER F.

MAZO, JANE LYNCY D.

June 2023

APPROVAL SHEET

In partial fulfillment of the requirements for the Degree of Bachelor of Science in Computer Engineering, this Design Project entitled **Ensuring Safety for Kindergarten Students: A Real-Time IoT-Based Alarm and Monitoring System at Manalupang San Vicente Elementary School** prepared and submitted by **Jan Miller Jaro, Jane Lyncy Mazo, Lourydem Dawis** is hereby accepted.

Engr. ANACIETO MERCADO
Adviser

PANEL OF EXAMINERS

Approved by the Committee on Oral Examination panel with a grade of__________.

Dr. JEN ALDWAYNE DELMO
Chairman

ENGR. MARK JOHN FEL T. RAYOS
Member

Dr. JEFFREY S. SARMIENTO
Member

Accepted as partial fulfillment of the requirements for the Degree of Bachelor of Science in Computer Engineering

DR. REYNATO A. GAMBOA
Dean, CEAFA

Date: ________________

DEDICATION

The proponents sincerely devote their hard effort to this project's study to the following:

To Lord God Almighty;

To their Parents;

To their Advisers;

To their Beneficiary;

To their Friends;

To their Loved Ones; and

To their Dean and Department Professors

ACKNOWLEDGEMENT

The completion of this project would not be possible without the continuous support and help of the family, friends, and peers of the researchers. Their contributions and support in conducting the research are sincerely appreciated. Moreover, the proponents would like to express their deep gratitude to the following persons:

Giving thanks to the Lord, the greatest Engineer of all, for being kind and providing strength to the developers, as well as thanks for the tremendous power that inspired us to do so.

To all the advisers, professors, deans, and panelists who selflessly offered their time and effort to teach, lead, and mentor us throughout the research process and in general.

To all studies and parents, who guided and shared their ideas in developing and validating this project.

Nevertheless, each member of the group showed their abilities in doing this study, sharing information and fun that lifted everyone's spirits despite the strain.

LBD

JMFJ

JLDM

ABSTRACT

Kindergarten is important when it comes to a child's education, which is why many parents are excited when their kid’s entering kindergarten. But sometimes other parents are worried about the safety of their kids in the classroom. School is their second home where children can learn a lot of important things. But their classroom is also a place where possible hazards might happen, especially when their children are vulnerable, curious, and sometimes unsupervised, which can lead to certain accidents. The importance of having assistance, especially for teachers, can really increase security.

In this research, it created an IOT-based alarm and real-time monitoring system that detects possible accidents based on safety hazards present in the kindergarten classroom at Manalupang San Vicente Elementary School. This created and evaluated three models in order to find the best model to solve this problem. With the use of YoloV7 and instance segmentation, it creates a model with 97% accuracy; with mask R-CNN, it reaches 94%; and with 2D pose estimation, it reaches 89%. Therefore, it used the YoloV7 instance segmentation model, then developed software and integrated it into hardware systems using the said model, which effectively detected safety hazards in the kindergarten environment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATION

CHAPTER I

INTRODUCTION

Introduction

This chapter outlines the background, objectives, scope, and delimitation, and importance of the Study.

1.1 Background of the Study

Kindergarten is a crucial milestone for a child's education, but for parents, it brings a mix of excitement and concerns regarding their child's safety. While school is a place where children can learn valuable lessons, it also poses potential risks, particularly for vulnerable and unsupervised young ones. Ensuring risk assessment and management is a vital responsibility for teachers to fulfill their duty of care. However, supervising a group of 10 or more adventurous children can be extremely challenging. Typically, a kindergarten class consists of around 20 students and a single teacher (Samuels, 2017).

In a tragic incident involving Eloisa's family, their child lost life due to an electric shock. Eloisa briefly left the child unattended in the presence of a metal object, which was connected to a socket, resulting in the unfortunate accident and fatality (Antonio, 2021). Young children are naturally curious and unaware of the dangers posed by unfamiliar objects or situations, making them particularly vulnerable.

Furthermore, Kitamura et al. (2020) conducted a study to address the issue of decreasing the number of children being admitted to emergency rooms due to accidents, particularly falls. The research focused on environmental modifications as a preventive measure, as relying solely on human supervision for injury prevention has proven ineffective. This utilized a posture database to simulate and visually understand children's daily climbing behaviors. Predicting children's behavior is crucial for improving the environment. It aims to further advance this approach by implementing it in various living spaces such as homes and kindergartens to gather more 3D posture data (Kitamura et al., 2020).

Ohnuki, Abe, and Suganuma (2020) proposed a method to detect pre-accident situations in infants or preschool children within a room and alert nearby adults. The aim is to prevent unforeseen accidents involving young children, as these incidents often require emergency medical assistance. The proposed method utilizes a video game to recognize the positions, postures, and movements of individual infants in the room. By referencing a room model, the system determines the relationship between the child and specified objects or areas, detecting potential danger. This approach goes beyond monitoring the child's posture and movements and considers their proximity to nearby objects. Preliminary experiments validated the system's ability to recognize the states of infants from a video game. The researchers plan to develop a prototype system and assess its effectiveness in detecting pre-accident situations among infants in a real room setting (Ohnuki et al., 2020).

Due to their extreme vulnerability to accidents, children need to be protected at all costs. In addition, parents and guardians are frequently among those who neglect children who are under their care. Children who are the victims of this neglect may suffer permanent harm, and in certain cases, their lives may be lost. Therefore, having an intelligent AI that can aid parents can keep kids safe from potential dangers. To assist and keep more kids from getting wounded, the researcher hopes to develop this study into a camera product that can be used at home, school, or even while driving. Further, this improve the function of this product by increasing the three behaviors that can lead to unforeseen accidents.

The previous research studies emphasized the importance of monitoring and detecting hazards to ensure children's safety. However, their focus is primarily on home environments, leaving a gap in detection capabilities for other settings such as classrooms. To address this, it proposed a system that monitors and detects hazardous situations based on human behavior and position, specifically in a classroom setting. The system works by alerting the teacher whenever it identifies potential accidents, allowing for immediate intervention to prevent injuries. Even when teachers are away, it can still monitor their students in real-time through IoT-enabled video streaming from the camera's perspective.

This aims to develop a system that can effectively monitor and detect possible accidents caused by safety hazards. It specifically targets children aged 5-6 years, which aligns with the typical age range for kindergarten according to MOM NEWS DAILY (2021). The system utilizes image processing to provide a visual perspective and employs a neural network with a trained algorithm and efficient datasets to analyze safety hazard situations.

A single Stream Camera is installed in the upper part of the classroom, connected to a mini-PC that houses the model. The proposed system's primary purpose is to intervene in potential dangers and supervise children. By alerting teachers about impending risks, it has the potential to save lives, particularly those of vulnerable children who may be unaware of the dangers around.

1.2 Objectives of the Study

The main objective is to create an IOT-based Alarm and Real-Time Monitoring System that detects possible accidents based on safety hazards present in the kindergarten classroom at Manalupang San Vicente Elementary School. The specific objectives are:

A. To gather an image dataset of various safety hazards found in the kindergarten environment such as climbing, running, and tampering with outlet.

B. To fine-tune and test various pretrained models then implement the most efficient model for safety hazard detection inside the classroom.

C. To embed the image detection model into the software and hardware systems of the device, which includes a stream camera, speaker, and mini-PC, designed specifically for the kindergarten classroom.

D. Evaluate the system's performance using the performance measures specifically accuracy and false positive rate.

1.3 Scope and Delimitation of the Study

This study focused on developing an IoT-based system that monitors and detects possible accidents based on safety hazards present at the kindergarten classroom of Manalupang San Vicente Elementary School. The camera is placed in the center of the upper front of the classroom, providing a wider view of the classroom.

The system is designed to detect safety hazards, which are the most common occurrences in a kindergarten classroom environment. Specifically, it can detect children climbing on chairs and tables, as well as running or tampering with electrical outlets. Depending on the type of hazard detected, the system triggers an appropriate alarm to alert the teacher.

Real-time viewing of the classroom is available only to the teacher, who can access the live stream to monitor the situation and act in case of an emergency. Access to live stream viewing is limited to ensure privacy and security and requires a reliable internet connection to maximize its potential. The device does not be able to maximize its features if there is no internet connection, as it is needed by the TeamViewer software. This software is used to access the mini-PC, and from there, the user can watch the real-time monitoring system.

The developed model is designed to operate effectively within the specific location where the dataset is collected. To utilize the model in different locations, it is necessary to acquire a new dataset, conduct training, and generate a fresh model for integration within the software. The model's performance is contingent upon the characteristics of the camera used during data collection, which had a limited field of view and required precise positioning for optimal monitoring. It is important to note that the current model exclusively detects safety hazards such as climbing, running, and tampering. In the event of new safety hazard classifications, it must be added to the dataset and included in the training process to enable the detection of these additional hazards. Lastly, the device is reliant on a stable electric supply, and without electricity at the deployment location, the model is non-functional.

1.4 Importance of the Study

The results of this study do not only benefit the researchers but also the parents, kindergarten teachers, and students of Manalupang San Vicente Elementary School. It provides parents and family members with the assurance that their kindergarten child is safe and well supervised while at school. Moreover, kindergarten teachers benefit from the study as it helps to monitor their students more conveniently, receive alerts if a child approaches a hazardous object, and have access to real-time viewing of the classroom, enabling them to take prompt action. Additionally, kindergarten students have peace of mind being monitored by their teacher and can easily receive help if a harmful situation arises.

Furthermore, this study provides baseline data for future research studies on monitoring the safety and security of children in kindergarten classrooms. This could also inspire future studies on the Internet of Things and instance segmentation. The findings of this study may have implications beyond Manalupang San Vicente Elementary School, as the results may be relevant to other kindergarten classrooms across the world facing similar challenges.

Definition of Terms

2D Pose Estimation: It involves estimating the positions of specific keypoints in a 2D space within an image or video frame. The model calculates X and Y coordinates for each key point.

AI Model: It refers to a computer model or program specifically designed for tasks related to artificial intelligence, such as image recognition or natural language processing.

Instance Segmentation: It's a type of image segmentation that focuses on identifying individual instances of objects and drawing boundaries around it.

Internet of Things (IoT): It refers to a network of connected devices, machines, objects, or even living beings (like animals or people) that can exchange data with each other without needing direct human or computer interaction.

Mask R-CNN: It's an advanced method for segmenting images using a type of neural network called Convolutional Neural Network (CNN). It can recognize objects in an image and generate precise segmentation masks for each object.

Model Architecture: It refers to the structure or design of a machine learning model, which includes the choice of algorithms and frameworks used to build it.

Neural Network: It's a technique used in artificial intelligence that enables computers to process information in a way inspired by the human brain.

OpenCV: It's an open-source software library used for computer vision and machine learning tasks. OpenCV provides a standardized framework for developing applications related to artificial intelligence, particularly in the field of computer vision.

Python: It's a high-level programming language that is easy to read and write. Python is often used for rapid application development and scripting, as it offers built-in data structures, dynamic typing, and dynamic binding.

YOLOv7: It's a type of neural network that can simultaneously predict bounding boxes (which define the location of objects) and their corresponding class probabilities (identifying what the objects are). This differs from previous object detection methods that used separate classifiers.

CHAPTER II

LITERATURE REVIEW

This chapter presents the conceptual literature, related studies, and conceptual framework respectively.

2.1 Conceptual Literature Hazards at Classroom

Falls are particularly likely to result in a serious injury. In fact, falls from height are the number one cause of fatal injury in UK workplaces. In a classroom environment, teaching staff or pupils using classroom furniture such as desks or swivel chairs to reach something up high can result in significant injury. Naturally, in a classroom environment, electrical equipment is likely to be used a lot, and with vulnerable people under your care, the need for good electrical safety is high. Young children are prone to slips, trips, and falls which is why proactive prevention is essential (Carracedo, R. (2021).

Extensions and loose wiring can also be a child hazard if the kid decides to play with them. Any damaged wiring can create sparks and be potential fire hazards if not maintained correctly. Children might also want to stand and hold onto chairs, tables, televisions, and other items that could tip over and fall on them (Barera, 2016).

In 2022 there were at least 96 incidents of gunfire on school grounds, resulting in 40 deaths and 78 injuries nationally (“EVERYTOWN”, 2022).

Theories Supporting the Problem

Finding the root of this problem, for Morrongiello B. (2018), proximity is particularly important for the safety of younger children under 6 years of age because it often do unpredictable things which increase exposure to and interactions with injury hazards.

Further, for Morrongiello & McArthur, at an uncertain time when children acquire new developmental milestones, which often occur unexpectedly for parents, injury rates show temporary peaks. Therefore, when children behave unpredictably, and parents have not had sufficient time to adjust the level of supervision children most frequently get injured.

Impact of Supervision

Different sources show the significant impact of supervision. Another popular testing approach to study supervision which was narrated by Morrongiello & McArthur(2018) involves the use of contrived hazards which appear real but that have been modified to pose no real risk of injury in laboratory settings. These methods yielded substantial insights regarding links between supervision and child injury risk. Furthermore, supervision is a strategy that has been shown to achieve this goal. Significantly, research has shown that children had a history of more medically attended injuries when parents reported reduced supervision but not when parents reported high levels of supervision. Thus, close supervision can counteract the elevated risk of injury typically found for temperamentally difficult children. Further, Wanyingo(2018) states that active supervision of children yields constant pay of attention to the children. Thus, gives the parents a possible chance to reach if the children are attempting to do something which can cause an injury and possibly worse. Further, parents can protect their children from injury when awareness occurs.

Concluding all the observations and results of various studies, improvement in supervision is a must for preventing possible danger or injuries the children may face, specifically those below 6 years old. Further, the defined possible hazards may face mainly involves hand reaching and, in this case, increasing the possible accuracy of the system which utilize hand movement for better detection of the hazard situation.

Safety Hazard

Safety hazards include unsafe working conditions that can cause injury, illness, or death. It consists of Spills on floors or tripping hazards such as blocked aisles or cords running across the floor, working from heights, electrical hazards like frayed cords, missing ground pins, and improper wiring (Escano, 2022).

Real-time Image Processing

Many real-world devices or products, such as mobile phones, digital still/video/cell-phone cameras, portable media players, personal digital assistants, high-definition television, video surveillance systems, industrial visual inspection systems, medical imaging devices, vision- assisted intelligent robots, spectral imaging systems, and many other embedded image or video processing systems, rely on real-time processing (Kehtarnavaz, 2020).

Open CV

OpenCV (Open Source Computer Vision Library) is a huge open computer vision, machine learning, and image processing library. It is compatible with a wide range of programming languages, such as Python, C++, and Java. It can analyze images and videos to identify objects, faces, and even human handwriting. When it is combined with other libraries, such as Numpy, a highly optimized library for numerical operations, the number of weapons in your arsenal grows, and whatever operation that can be done in Numpy can be combined with OpenCV (GeeksforGeeks, 2021).

Instance Segmentation

Instance segmentation is a computer vision task for detecting and localizing an object in an image. Instance segmentation is a natural sequence of semantic segmentation, and it is also one of the biggest challenges compared to other segmentation techniques. The goal of instance segmentation is to get a view of objects of the same class divided into different instances (Kurtanović, 2021).

Fine Tuning

To get the desired results, this technique involves fine-tuning a few network levels. To make the pre-trained model more applicable to the given problem, some representations of the model have been slightly altered. By doing this, the neural network can be trained without having to define its structure first (Torres, 2023).

Mini PC

A mini PC is a compact computer that has become popular among developers due to its small size, low power consumption, and powerful components. Mini PCs offer cost-effective solutions on a budget and are highly portable, allowing to work from anywhere. Additionally, mini PCs consume less energy than regular-sized computers, making them a more sustainable and cost-effective option in the long run. With powerful processors, high-speed RAM, and solid-state drives, mini PCs are capable of handling complex development tasks such as coding, compiling, and testing software. Overall, mini PCs are an ideal option for developers who require a portable and flexible computing solution without compromising computing power (GeekomPC, 2021)

Cloud Server

A cloud server is a pooled, centralized server resource that is hosted and delivered over a network, typically the Internet, and accessed on demand by multiple users. Cloud servers can perform all the same functions of a traditional physical server, delivering processing power, storage, and applications (“VMware”, 2022).

Internet of Things

The internet of things, or IoT, is a system of interrelated computing devices, mechanical and digital machines, objects, animals, or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to- computer interaction(Gillis, 2022).

Stream Camera

Stream camera features automatic HD light correction, which causes the camera to modify image quality and adapt to various lighting situations so that the video appears bright and clear even while it is being recorded in dim light (Adorama Learning Center, 2021).

2.2 Related Literature

This paper focuses on improving the detection accuracy of hazardous material vehicles using the YOLOv5 method. The current detection technology faces challenges related to computational effort and accuracy. To address these issues, the proposed method incorporates an attention module in the YOLOv5 backbone and neck networks to extract better features by assigning different weights to different parts of the feature map. The SPPCSPC layer replaces the SPPF layer to enhance feature fusion across different-sized feature maps. Furthermore, the SIoU loss function is introduced to enhance bounding box regression and localization accuracy. Experimental results on the dataset demonstrate that the improved model significantly enhances the detection accuracy of hazardous chemical vehicles compared to the original model. This approach holds great significance for real-time detection during the driving process, allowing for timely recognition of traffic accidents and effective emergency responses, thereby preventing casualties and property damage. In summary, this research presents an improved algorithm based on YOLOv5 for detecting hazardous material vehicles. It incorporates an attention mechanism, utilizes the SPPCSPC layer for feature fusion, and introduces the SIoU loss function for accurate localization. The experimental results validate the effectiveness of the proposed method, showcasing improved detection performance for hazardous chemical vehicles, including smaller vehicles in complex scenes (Zhu et al. 2023).

The study addresses the need for regular inspection and evaluation of potential safety hazards (PSHs) along high-speed railroads. The current visual inspection approach is time-consuming and may pose safety concerns for inspectors, especially in remote areas. To complement visual inspections and mitigate safety concerns, the study develops an automatic PSH detection framework called YOLARC (You Only Look at Railroad Coefficients) using unmanned aerial vehicle (UAV) imagery. YOLARC incorporates a new backbone with multiple receptive fields to enhance the multi-scale representation capability and semantic information in the feature space. It utilizes a lightweight feature pyramid network (FPN) with a multi-scale pyramidal architecture and a Protonet with a residual structure to accurately predict track areas and PSHs. Additionally, a hazard level evaluation (HLE) method is developed to quantify the level of hazard by calculating the distance between identified PSHs and the track. Experiments conducted on UAV imagery datasets demonstrate that YOLARC efficiently converts UAV images into valuable information with high detection rates and processing speeds. It outperforms other object detection and instance segmentation models, achieving the best average precision (AP) of 53.98%. YOLARC also provides track area segmentation and accomplishes hazard level evaluation for detected PSHs. Compared to other models, YOLARC offers faster processing speeds while maintaining high detection accuracy. The study represents the first attempt to apply UAV images and instance segmentation for PSH inspections along high-speed railroads. The proposed system demonstrates promising performance in terms of real-time speed, accuracy, and robustness in complex railroad environments. Future work aims to develop a customized software package with low RAM consumption, computational cost, and high accuracy, suitable for implementation on edge computing platforms mounted on UAVs for field practices (Wu et al. 2022).

The research focuses on the importance of safety helmet usage in industries such as construction, power, and manufacturing. Employees often remove their helmets, leading to increased risks and injuries. Conventional methods of helmet identification lack accuracy and efficiency. Therefore, the study proposes a deep learning approach using the Kaggle dataset. A total of 5000 images are collected, with 1000 images used for testing. Region-based Convolutional Neural Network (RCNN) and You Only Look Once (YOLO) models are compared for helmet detection. YOLO achieves the highest mean Average Precision (mAP) success rate of 97.12%. Real-time implementation of deep learning is crucial for identifying safety violations and reducing manual monitoring efforts. The study concludes that YOLO is the most effective model for helmet detection and suggests future work on designing an alerting system for supervisors to address non-compliance (N.K Anushkannan et. al. 2022).

This research conducted by Arief et al. (2020) addresses the issue of smoking in public facilities and proposes a more comprehensive solution. The current approach of using smoke detectors is more detective than preventive. The aim is to develop a system that not only triggers an alert upon detecting smoke but also recognizes cigarette objects to promote prevention. When smoke or a cigarette is detected, a public service announcement (PSA) is automatically displayed to educate individuals about the hazards of smoking. The system combines embedded systems and deep learning, continuously monitoring smoke levels from sensors and performing cigarette object detection using YOLO. The outputs are used to trigger an anti-smoking PSA and a buzzer. The system includes a Raspberry Pi 4B 4GB, smoke sensors, a camera module, a monitor, and a buzzer. The results demonstrate the system's ability to detect smoke and cigarettes in a semi-indoor environment, triggering alerts and displaying PSAs in a soft real-time manner.

Further, key conclusions from the research are as follows: 1) The optimal distance for detecting cigarette objects using YOLO is 2 meters with a 5MP camera and a resolution of 192x192. 2) The camera used may not be sufficient for distinguishing objects with similar shapes and characteristics. 3) The focus on speed compromises the accuracy and precision of object detection for similar objects beyond 2 meters. Improving accuracy can be achieved by adjusting input amounts and balancing the training dataset. 4) Smoke detection can be performed using the MQ-7 smoke sensor at an optimal distance of 0-15 cm. 5) The display of PSAs as educative warnings follows a pulsing model. Possible future developments suggested by the research include adding smoking gestures/behaviors detection, which is less affected by sensor distance and sensitivity, and employing transfer learning to improve the speed of real-time object detection (Arief et al. 2020).

The paper of Zhang &Zhou (2021) addresses safety detection in underground coal mines (UCM), which is hindered by challenging environmental factors such as dim light and dense dust. To overcome the lack of publicly available coal mine environment datasets, the authors constructed a simulated dataset called SDUCM-dataset. This dataset accurately simulates real coal mine scenes and fulfills the requirements of common safety detection tasks. The study introduces a novel safety detector called YOLOUCM, based on the YOLOv5 (You Only Look Once version 5) architecture. YOLOUCM is designed to efficiently detect potential safety hazards in UCM with higher efficiency than most one-stage object detectors, while maintaining comparable accuracy to two-stage object detectors. The model incorporates various techniques such as Vision Transformer, Merge Non-Maximum Suppression, and Meta-AconC to enhance performance without sacrificing detection speed. Experiments are conducted to evaluate the proposed method for pedestrian safety detection in UCM using the SDUCM-dataset.

The results of the research conducted by Zhang &Zhou (2021) demonstrate that the dataset effectively supports training of YOLO-UCM, achieving a delicate balance between accuracy and speed in UCM environments. Ablation experiments showcase the advantages of YOLO-UCM over other commonly used object detection algorithms in UCM. The authors acknowledge the limitations of their work, including the insufficient number of images in the dataset and the categories of labels. They propose future work to address these limitations by incorporating Generative Adversarial Networks (GANs) to construct a more comprehensive and effective dataset with improved label attributes. Additionally, they consider adopting knowledge distillation to create a smaller parameter model with similar performance, enabling the proposed model to be applied in actual UCM scenarios.

The study focuses on the factors contributing to fatal and injury-causing incidents in India, with potholes being a major concern. Excessive speed, driver distraction, failure to wear safety equipment, and road hazards like potholes are identified as primary causes of accidents. The rising number of vehicles on Indian roads, with over 295 million registered vehicles, further exacerbates the pothole issue. Existing methods for pothole detection using sensors or specialized hardware are either expensive or challenging to implement widely. To address this, the proposed work introduces a new method that utilizes a smartphone's camera and location data to classify and count potholes. The YOLO v7 algorithm is employed for pothole categorization due to its speed and accuracy. The algorithm has been found effective for detecting potholes of various sizes and shapes. The research involves extensive testing and training using a dataset of photographs, employing the YOLO technique for classification, and utilizing the Google API and accelerometer for pothole counting. In summary, the study recognizes the significance of potholes as a major contributor to accidents in India. It proposes a novel approach that leverages a smartphone's camera and location to classify and count potholes. The YOLO v7 algorithm is employed for accurate pothole detection, and the Google API and accelerometer are used for counting. The research aims to provide a more affordable and accessible method for pothole detection and monitoring (Reddy & V, 2022).

This paper addresses the security management challenges in substations, specifically the timely detection of fire hazards, unauthorized persons, and engineering vehicles. Traditional methods require additional hardware installation, which is costly and limited to detecting specific security threats. To overcome these limitations, the paper proposes a substation early-warning system using video surveillance information and the YOLO-v5 deep neural network for real-time object detection. The system utilizes a regression model with a deep convolutional neural network that combines the Backbone and PANet structures. A substation image dataset is augmented to include various weather conditions and labeled to indicate multiple hazards. Through stochastic gradient descent and minimizing the YOLO-v5 integrated loss functions, the deep neural network is trained to identify early-stage fires/smokes, unauthorized objects, and abnormal positions of vehicles in substations. Experimental results demonstrate that the proposed method can automatically identify multiple safety hazards in substations in real-time, thereby enhancing the overall security level. It effectively detects potential risks such as smoke, open fires, abnormal person intrusion, and dangerous proximity of construction vehicles. By enabling timely alarms and prompt actions, the system improves the safety management of substations and offers promising applications in the field (Xiao et. al, 2022).

This paper focuses on ensuring the safety of substation personnel by detecting and monitoring abnormal behaviors, such as not wearing safety helmets, entering dangerous areas, or smoking, which are prohibited by security regulations. Leveraging the availability of video monitoring equipment in substations, the paper presents a security monitoring system that utilizes automatic image object detection methods. The system takes video monitoring as input and employs a regression deep convolutional neural network for predicting the probability of abnormal behaviors. Based on the YOLO-V5 algorithm, the deep neural network is trained using a multi-part loss function that considers locating, sizing, classifying, and the probability of abnormal behaviors. Experimental results demonstrate that the proposed method can effectively and confidently monitor abnormal behaviors of individuals in substations, including locating each person and identifying whether they are wearing helmets. These findings validate the potential application of the proposed method in complex substation environments (Xiao, 2022).

In summary, this paper develops an automatic substation abnormal behavior monitoring system based on a deep convolutional neural network and the YOLO-V5 algorithm. The system can accurately predict the probability of abnormal safety behaviors in real-time. By minimizing a multi-objective loss function, the neural network is trained to provide monitoring, timely early warning, and prediction of abnormal behaviors in substations. The experimental results confirm the system's ability to automatically monitor personnel's abnormal behaviors, determine helmet usage, and effectively operate in the complex background of substations, showcasing its promising application potential (Xiao, 2022).

This study addresses the need for safety and monitoring in exam halls by proposing an Abnormal Behavior Detection Technique. Currently, exams are monitored manually or through video surveillance, which can be time-consuming and prone to inaccuracies. The study aims to develop a method that not only identifies hazardous actions but also helps reduce them. The proposed technique involves using interdisciplinary approaches for outlier identification and classification. It focuses on strategies for detecting abnormal behaviors in exam halls and discusses the advantages and disadvantages of different approaches. The goal is to identify any new unusual behavior exhibited by students during exams. The recommended method begins by locating the region of interest (ROI) for each student. A vision-based YOLO (You Only Look Once) detection method is employed to identify the student's face, upper body, and head. The system continuously monitors the behavior of students in the exam hall and provides messages based on their normal or abnormal movements.For the system to be successful, the vision algorithms need to work in all conditions. The proposed method has the potential to help both examiners and academic institutions identify unethical practices in the exam hall. The accuracy of the proposed system was reported to be 87.20 percent. In summary, this study proposes an Abnormal Behavior Detection Technique for monitoring exam halls. It suggests using vision-based algorithms to detect abnormal behaviors and provide timely notifications. The method has the potential to improve safety and ethics in exam environments, benefiting both examiners and academic institutions (Padhiyar et. al. 2023).

This study introduces an UAV image-based intelligent inspection method for detecting safety hazards along high-speed railways. The proposed method utilizes an All-in-One YOLO architecture (AOYnet) that combines semantic segmentation and object detection in parallel tasks, addressing the inefficiency of single-task detectors. The method achieves high performance on the Railway Surrounding Environment Dataset, outperforming existing methods in terms of mean Average Precision (mAP) for detected objects and Intersection over Union (IoU) for segmented objects. The proposed method demonstrates efficiency and feasibility in detecting environmental hazards using UAV images. To enhance detection accuracy, a specific UAV dataset for railway surrounding environments is established. The end-to-end AOYnet model handles multiple tasks relevant to railway scenes and shows improved hazard detection performance while considering complex environments. However, challenges remain, as not all hidden hazards can be accurately identified in the complex railway surroundings. Hidden hazards may only be recognized as such once they are discovered. In conclusion, this study proposes an UAV image-based intelligent inspection method for detecting environmental hazards along high-speed railways, achieving superior results and offering efficiency and feasibility.

Addressing hidden hazards and improving detection accuracy in railway surroundings require further research (Chen et. al,2022).

This study presents a novel insulator defect detection method using improved YOLOv7 and a multi-UAV collaborative system to overcome limitations of single UAV-based systems. A comprehensive dataset including various types of defects is constructed, and comparative analyses show that the proposed YOLOv7-C3C2 and YOLOv7-C3C2-GAM models outperform the baseline YOLOv7 model. Experimental results validate the method's effectiveness in enhancing accuracy and reliability. The multi-UAV system demonstrates superior flexibility and detection speed compared to single UAV systems. The proposed YOLOv7-C3C2-GAM model leveraging modules and attention mechanisms, improves feature extraction and reduces network complexity. Challenges remain in real-time communication, path planning, and detection performance under extreme weather conditions and for small targets. Future enhancements can leverage camera advancements for better defect detection. Overall, this study provides a theoretical and technical framework for insulator defect detection, reducing labor intensity and meeting inspection requirements. Further improvements are possible, expanding the method's potential applications beyond power equipment defect detection (Chang et al. 2023).

The paper addresses the issue of incompatibility between autonomous vehicles and vulnerable road users in the traffic system, particularly at pedestrian crosswalks in urban areas. The authors propose a warning system that aims to reduce accidents by providing simultaneous warnings to drivers, disabled individuals, and pedestrians with phone addiction. The objective of the study is to automatically detect pedestrian crosswalks using both vehicle and pedestrian perspectives.The authors employ two network models, Faster R-CNN (R101-FPN and X101-FPN) and YOLOv7, for the analysis of a dataset collected by the authors. The accuracy comparison between the models reveals that YOLOv7 achieves an accuracy of 98.6%, while Faster R-CNN achieves 98.29%. YOLOv7 demonstrates better prediction results than Faster R-CNN for different types of pedestrian crossings, although the overall results are quite similar (Kaya et al. 2023).

The study focuses on pedestrian crosswalks as critical areas in traffic safety and the autonomous process. The detection process considers the perspectives of both vehicles and pedestrians, emphasizing the mutual autonomy required in the traffic network.The analytical results indicate that both proposed models perform above an acceptable level, with both achieving a detection accuracy of around 98%. YOLOv7 shows slightly better results in the comparison between the two models. Augmenting the dataset is suggested as a means to potentially increase accuracy. The effectiveness of the study could be further enhanced by integrating an embedded system. The authors are already working on a warning system and have the potential to develop the existing dataset over time and explore different detection models.In summary, the study focuses on the automatic detection of pedestrian crosswalks in urban areas, considering both vehicle and pedestrian perspectives. The proposed YOLOv7 and Faster R-CNN models exhibit high detection accuracy, with YOLOv7 showing slightly better results. Future directions include dataset augmentation, integration of an embedded system, and the potential development of new detection models. (Kaya et al. 2023)

The study introduces YOLOv7-sea, an improved object detection algorithm specifically designed for maritime search and rescue missions. The researchers address the challenges posed by the SeaDronesee dataset, which includes small targets and significant sea surface interference, making it difficult for conventional object detectors to perform effectively. To enhance detection capabilities, the researchers incorporate a prediction head to identify tiny-scale objects and people. Additionally, they integrate the Simple, Parameter-Free Attention Module (SimAM) to identify attention regions within the scene. Several strategies such as data augmentation, test time augmentation (TTA), and bundled box fusion (WBF) are employed to further enhance the proposed YOLOv7-sea algorithm. The evaluation on the ODv2 challenge dataset shows that YOLOv7-sea achieves an AP result of 59.00%, surpassing the baseline model (YOLOv7) by approximately 7%. These advancements contribute significantly to the field of object detection in maritime scenarios, particularly for analyzing images captured by maritime drones and improving search and rescue operations. The findings of this study offer valuable insights and guidance for developers and researchers in this domain (Zhao et al. 2023).

This work introduces a novel framework that combines Unreal Engine, Microsoft AirSim, and Robot Operating System (ROS) to control an Unmanned Aerial System (UAS) and perform real-time object detection in a dynamic sea simulation environment. The framework allows for easy manipulation of various simulation attributes such as lighting, camera settings, and environmental conditions. The primary objective was to optimize the UAS flight altitude and camera brightness for accurate detection of human victims. The study evaluates the performance of two pretrained YOLOv7 models: YOLOv7-SDS, retrained on the SeaDronesSee Dataset, and YOLOv7-COCO, originally trained on the Microsoft COCO Dataset. Through extensive manual testing and autonomous UAS missions, the framework achieved high accuracies of 97.8% for YOLOv7-SDS at an altitude of 8m with high camera brightness, and 93.79% for YOLOv7-COCO at a height of 2m from the sea-level with lower camera brightness (Poudel et al. 2023).

The developed framework shows potential for further research. While this study presents a limited number of experiments, the framework can be utilized by other researchers to conduct their own experiments and explore different object detection models and configurations. Additionally, the framework can be reverse engineered to automatically generate ground truth labels from the gaming engine, enabling real-time training of object detection models in diverse marine environments. This has the potential to revolutionize maritime computer vision. In summary, this work introduces a framework combining Unreal Engine, Microsoft AirSim, and ROS for controlling a UAS and performing real-time object detection in a sea simulation environment. The framework achieves high accuracies for detecting human victims using pretrained YOLOv7 models. Further experimentation and the potential for real-time training of object detection models are highlighted as future research directions. The framework's capabilities and potential contributions to maritime computer vision are emphasized (Poudel et al. 2023).

This paper addresses the urgent need for an automated system that can detect falls in low-light conditions, particularly for elderly individuals living alone with limited access to healthcare. Conventional vision-based systems for fall detection often fail in low illumination, making it crucial to develop a system that can protect vulnerable individuals. The proposed system introduces a novel vision-based fall detection approach that combines object tracking and image enhancement techniques. The method is divided into two main parts. First, a dual illumination estimation algorithm optimizes the captured frames to enhance visibility in suboptimal lighting conditions. Then, a deep-learning-based tracking framework, utilizing YOLOv7 for detection and the Deep SORT algorithm for tracking, is employed for fall detection. The effectiveness of fall detection in dark night environments with obstacles is evaluated using the Le2i fall and UR fall detection (URFD) datasets. The proposed method integrates dual illumination estimation with the YOLOv7 + Deep SORT tracking algorithm to enhance fall detection performance in dim or uneven lighting conditions. Additionally, exposure correction techniques are incorporated to improve fall detection in videos. The performance of the proposed method is evaluated using the Le2i FDD and UR-Fall datasets. For future experiments, the authors plan to implement a self-learning framework that can adapt to false alarms and improve the performance of existing fall detection systems by incorporating correct results. In summary, this paper presents a vision-based fall detection system that combines deep learning-based tracking with dual illumination estimation and exposure correction. The proposed method demonstrates improved fall detection performance in low-light conditions. The evaluation conducted on the Le2i FDD and UR-Fall datasets supports the effectiveness of the proposed approach (Zi et al. 2023).

This paper focuses on addressing the inefficiency, labor cost, and inaccuracies associated with the manual counting method used in the hemp duck breeding industry. It proposes the use of deep learning algorithms for real-time monitoring of dense hemp duck flocks to improve productivity and animal welfare. The study constructs a new large-scale hemp duck object detection dataset and introduces an improved attention mechanism YOLOv7 algorithm, CBAM-YOLOv7, along with SE-YOLOv7 and ECA-YOLOv7 for comparison. Experimental results show that CBAM-YOLOv7 achieves higher precision and slightly improved recall, mAP@0.5, and mAP@0.5:0.95 compared to SE-YOLOv7 and ECA-YOLOv7. The evaluation indices of CBAM-YOLOv7 outperform the other algorithms. A comparison between two labeling methods, full-body frame labeling and head-only labeling, reveals that the full-body labeling method provides better detection results, as the head-only labeling method leads to the loss of important feature information. The proposed intelligent hemp duck counting method is deemed feasible and has the potential to promote the development of reliable automated duck counting. The study contributes a large-scale dataset for estimating hemp duck counts, facilitating visual research in the poultry field. The addition of CBAM modules to the YOLOv7 network structure improves detection accuracy without significant computational pressure. Future research directions include further optimizing the proposed algorithm's network structure and deploying it in the hardware environment used in field farming. In summary, this study proposes an intelligent method for counting dense hemp duck flocks using deep learning algorithms. The CBAM-YOLOv7 algorithm achieves improved detection accuracy, while full-body frame labeling proves more effective than head-only labeling. The research provides a valuable dataset and discusses the potential for future optimization and hardware deployment (Jiang et al. 2022).

This paper focuses on predicting the diagnosis of musculoskeletal disorders among sewing machine operators in the garment industry. Work-related injuries and disorders in this industry are often overlooked, leading to physical health damage, work time loss, and decreased productivity. The study records working videos of 20 participants (10 healthy and 10 unhealthy) from both sides, left and right. The OpenPose algorithm is employed for posture evaluation, estimating 2D human pose and extracting joint angles. These angles are normalized and used to build a classification model using the KNN Classifier. By implementing stratified k-fold cross-validation with 10 folds, the proposed approach achieves an accuracy of 91.3% in diagnosing musculoskeletal disorders among sewing machine operators. The research emphasizes the significance of addressing musculoskeletal disorders in the garment industry and highlights the applicability of advanced machine learning algorithms in this field. Managing operator workload based on their physical health becomes easier with the classification obtained. While the research provides a practical solution to a prominent problem in the garment industry, the authors acknowledge that further work can be done. Future research could explore additional experiments, testing, and models to gain deeper insights into the prevalence of musculoskeletal disorders. Suggestions for future work include applying other advanced machine learning algorithms to improve accuracy and developing a skill matrix for the production department based on workers' skills and risk classifications. Overall, the paper presents a systematic approach to predict and diagnose musculoskeletal disorders among sewing machine operators, highlighting the potential for further research and practical implementation in the industry (Chen et al. 2021).

This study explores the application of neural architecture search, specifically neuroevolution inspired by biological evolution, for designing efficient 2D human pose networks. The researchers propose a novel weight transfer scheme to accelerate the neuroevolution process effectively. The results demonstrate that the networks generated through this method are both more efficient and more accurate compared to existing hand-designed networks. Notably, the evolved networks achieve higher accuracy in processing higher-resolution images while requiring less computational resources than previous networks operating at lower resolutions. The base network, EvoPose2D-S, achieved comparable accuracy to SimpleBaseline, but with a 50% increase in speed and a 12.7x reduction in file size. The largest network, EvoPose2D-L, achieved state-of-the-art accuracy on the Microsoft COCO Keypoints benchmark while being 4.3x smaller than its nearest competitor and maintaining similar inference speed. The study concludes by highlighting the proposed weight transfer scheme's simplicity and effectiveness, as well as the successful utilization of large-batch training to accelerate the neuroevolution of efficient 2D human pose networks. This work represents the first application of neuroevolution to 2D human pose estimation, with additional experiments demonstrating that 2D human pose networks can be trained with batch sizes up to 2048 without sacrificing accuracy. By leveraging large-batch training and the proposed weight transfer scheme, the researchers evolved a lightweight 2D human pose network design suitable for mobile deployment. The EvoPose2D network, designed using neuroevolution, outperformed existing models in the literature when scaled to higher input resolutions while maintaining lower computational costs (Mcnally et al., 2018).

The study introduces an online method for efficiently detecting and tracking the 2D pose of multiple individuals in video sequences. It builds upon the Part Affinity Field (PAF) representation designed for static images and proposes an architecture that can encode and predict Spatio-Temporal Affinity Fields (STAF) across the video frames. The approach incorporates a novel temporal topology that is cross-linked across limbs to effectively handle various body motions. It also adopts a recurrent structure where the network uses STAF heatmaps from previous frames to estimate those for the current frame. The proposed method is designed for real-time reactive systems, achieving both high accuracy and fast performance that is independent of the number of people in the scene and the input frame rate of the camera. It demonstrates competitive results on the PoseTrack benchmarks and shows stability in tracking accuracy at reduced frame rates. This makes it suitable for deployment on low-power embedded systems that may have limitations in running large networks at high frame rates while maintaining reasonable accuracy. The study suggests potential future enhancements, such as incorporating a re-identification module to handle individuals leaving and reappearing in the camera view, as well as exploring the benefits of detecting and triggering warm start at shot changes to improve pose estimation and tracking performance Hidalgo et al., 2018).

In their research, the authors presented a real-time approach for multi-person 2D pose estimation called OpenPose. The goal of this work was to enable machines to understand and interpret humans and their interactions visually. The proposed method utilized a nonparametric representation known as Part Affinity Fields (PAFs) to associate body parts with individuals in an image. This approach achieved high accuracy and real-time performance, regardless of the number of people present in the image. Previous approaches focused on refining PAFs and body part location estimation simultaneously throughout training stages. However, the authors demonstrated that refining only PAFs, instead of both PAFs and body part locations, significantly improved both runtime performance and accuracy. They also introduced a combined body and foot keypoint detector, leveraging an annotated foot dataset released to the public. This combined detector reduced inference time compared to running the detectors sequentially while maintaining the accuracy of each component. The paper concluded that real-time multi-person 2D pose estimation is a crucial aspect for machines to comprehend humans and their interactions visually. The authors emphasized the nonparametric representation of keypoint association, the joint learning of part detection and association, and the efficiency of the greedy parsing algorithm for producing high-quality body pose parses. They also highlighted the importance of PAF refinement and the benefits of combining body and foot estimation into a single model. The research led to the development of OpenPose, an open-source real-time system for multi-person 2D pose detection. This system includes detection of body, foot, hand, and facial keypoints. OpenPose has gained widespread usage in various research areas related to human analysis, such as human re-identification, retargeting, and Human-Computer Interaction. Additionally, OpenPose has been integrated into the OpenCV library. The authors also created and plan to release a foot keypoint dataset consisting of 15,000 foot keypoint instances (Simon et al., 2019).

The study introduces a multitask deep learning framework for the simultaneous estimation of 2D and 3D poses from still images, as well as the recognition of human actions from video sequences. Traditionally, these tasks have been treated as separate problems in literature. However, the researchers demonstrate that a single architecture can efficiently solve both tasks and achieve state-of-the-art performance. The study also highlights the advantages of end-to-end optimization over separate learning approaches, showing significantly improved accuracy. Notably, the proposed architecture enables seamless training with data from different categories. The effectiveness of the method is validated through experiments conducted on four datasets: MPII, Human3.6M, Penn Action, and NTU. The results underscore the capabilities of the multitask deep learning model for the targeted tasks. Specifically, the model predicts the 2D and 3D coordinates of body joints in RGB frames, which are then utilized for action recognition using both semantic and visual information. The weight and feature sharing within the model allow for efficient and superior performance compared to dedicated approaches. Overall, the study contributes to the advancement of joint 2D/3D pose estimation and action recognition using a single, multitask architecture. (Picardi et al., 2018).

The study presents an analysis and approach that applies vision-based pose estimation to predict the performance of amateur golf swings using edge processing. The proposed system utilizes both still photos and videos to identify key frames in the golf swing, allowing for feedback and improvement. It can assess various metrics such as posture, swing tempo, and swing consistency, and even predict the outcome of the swing through path projection. The computation and data processing are performed on a low-cost tensor processing unit (TPU) to ensure optimal performance of the video capturing system. The study also addresses limitations and inaccuracies in hardware and pose estimation by employing a Savitzky-Golay filter. The goal is to develop a marker less swing tracking analysis system that is affordable and compact in size. In conclusion, the study initially focused on classifying a golf swing based on a single frame but evolved into recognizing patterns of movements or poses. The researchers experimented with different methods to address the problem incrementally. It first determined the system's constraints by directly running PoseNet on the hardware. Then, it gathered a set of photos from down the line shots for activity classification. A Key Frame Selector model was created to identify frames of interest, and videos were collected at a driving range to generate additional data, which revealed the importance of data filtering and a deeper understanding of camera hardware. Finally, a KFS plotter was developed to predict swing outcomes based on keypoints (Zohdy et al., 2020).

In this research study, the researchers explore the challenge of deploying high-resolution human pose estimation models on resource-constrained edge devices. This identifies that the high-resolution branches of HRNet-based models are unnecessary for low-computation scenarios, leading to increased computational costs. To address this, it proposes a new architecture called LitePose, which is a single-branch model that achieves real-time multi-person pose estimation on edge devices while improving both efficiency and performance. The authors conduct gradual shrinking experiments to investigate the redundancy of high-resolution branches in low-computation scenarios. Based on their findings, they develop LitePose by removing these redundant branches. Additionally, they introduce two approaches to enhance LitePose's capacity: fusion deconv head and large kernel conv. Their experiments on mobile platforms demonstrate that LitePose significantly reduces latency by up to 5.0 times without compromising performance. This advancement pushes the boundaries of real-time multi-person pose estimation on edge devices. In conclusion, the research focuses on efficient architecture design for multi-person pose estimation on edge devices. The authors present LitePose as an effective solution, which combines the advantages of both single-branch and multi-branch architectures. Through extensive experimentation, they validate the effectiveness and robustness of LitePose, opening possibilities for real-time human pose estimation in edge applications (Wang et al., 2022).

The adapted multi-person poses estimation architecture for edge devices, based on the bottom-up approach of OpenPose, was investigated in the study. The proposed network design and optimized post-processing code allowed the solution to achieve real-time performance on CPUs. On an Intel® NUC 6i7KYB mini PC, the solution achieved a frame rate of 28 frames per second (fps), while on a Core i7-6850K CPU, it achieved 26 fps. The network model used had 4.1 million parameters and a complexity of 9 billion floating-point operations (GFLOPs), which was only around 15% of the baseline 2-stage OpenPose, while maintaining a comparable quality. The study concluded that the accuracy versus network complexity ratio was significantly improved by utilizing a dilated MobileNet v1 feature extractor with depthwise separable convolutions and a lightweight refinement stage with residual connections. The network, named human-pose-estimation-0001, can be downloaded as part of the OpenVINO Toolkit, and its description is available in the Open Model Zoo repository. The solution demonstrated real-time performance on common CPUs and NUC mini PCs, with accuracy closely matching the baseline 2-stage network. The study also suggested additional techniques, such as quantization, pruning, and knowledge distillation, that could further enhance performance and accuracy (Osokin, 2018).

This paper addresses the importance of recognizing hand signals from Vulnerable Road Users (VRUs) such as bicyclists, scooter riders, e-scooter riders, and motorbike riders to enhance safety in Advanced Driver Assistance Systems (ADAS) and Autonomous Driving (AD). The objective is to reduce the probability of road accidents by enabling ADAS systems to understand VRUs' intentions and warn the car driver accordingly. The proposed system utilizes Convolutional Neural Networks (CNN) and Region Multi-Person Pose Estimation (RMPE) framework for human pose estimation to recognize VRUs' hand signals.The algorithm developed in this research can accurately identify all six hand signals of bicyclists (left, right, stop, give way, slow down, and road hazard), which has not been addressed in previous works. Additionally, the algorithm efficiently recognizes arm signals of scooter riders, e-scooter riders, and motorbike riders. The self-captured dataset for

VRUs Arm Signal Recognition is used for testing, along with the Cyclist Arm Signal Recognition (CASR) dataset for generalization. The proposed approach achieves state-of-the-art results in VRUs hand signal recognition with a frame rate of 10 FPS and 92% accuracy.This research focuses on real-time VRUs hand signal recognition, introducing a novel algorithm that can recognize a comprehensive set of hand signals for bicyclists, including those not addressed in previous works. It extends the recognition to arm signals of other VRUs such as scooter riders, e-scooter riders, and motorbike riders. The proposed algorithm is tested on self-captured data and the CASR dataset. In the future, the work can be expanded to include VRUs' orientation. Overall, this research contributes to improving the safety of ADAS and AD systems by accurately recognizing VRUs' hand signals in real-time scenarios. (Ashtekar et al. 2021)

This paper addresses the concern of falls as a significant health threat to the elderly population and proposes a fall detection model based on the OpenPose human posture estimation algorithm. The model utilizes machine vision techniques to detect falls and improve the accuracy of detection. The approach combines OpenPose's human key point detection with the SSD-MobileNet object detection framework to eliminate non-human key points detected by OpenPose, thereby reducing false detections, and enhancing algorithm robustness in complex environments. The model extracts feature from human joint points and employs the SVDD classification algorithm for classification. Experimental results demonstrate that the proposed method effectively detects falls with an accuracy rate exceeding 93%. The presented research introduces a fall detection model based on the OpenPose algorithm. By leveraging human key point detection in OpenPose and addressing the lack of fall-related data in the COCO dataset, the model incorporates migration learning to enhance accuracy. Additionally, the combination of SSD-MobileNet framework helps filter out non-human key points identified by OpenPose, leading to the extraction of key point features and subsequent classification using the SVDD algorithm. Experimental findings confirm that this approach reduces false detections and improves robustness in complex environments. (Sun & Wang, 2020)

This paper focuses on the importance of estrus detection in cattle for effective cattle breeding management in precision livestock farming. Accurate estrus detection enables timely artificial insemination, which greatly impacts the productivity of livestock farms. The authors propose an estrus detection approach that utilizes video inputs to track and identify cattle mating postures individually. To achieve precise identification and obtain individual cattle information, the paper introduces a vital step of segmenting each cattle from the background. An instance segmentation approach based on a Mask R-CNN deep learning framework is proposed to solve pixel-level segmentation masks for cattle in outer ranch environments. The segmentation of individual cattle is performed first to detect mounting behaviors, followed by a lightweight tracking algorithm as a post-processing step, which is an innovative aspect of this study. The training data for the proposed approach were collected by installing surveillance cameras at a livestock farm, while testing data consisted of various datasets from different camera placements. The experimental results demonstrate that the proposed approach achieves a detection accuracy of 95.5% in identifying estrus behaviors in cattle. In summary, this study presents an estrus detection system that utilizes deep learning-based object detection, robust cattle segmentation using Mask R-CNN, and a lightweight tracking algorithm. The proposed system performs well in complex ranch environments, even with a relatively small amount of training data. The achieved detection accuracy rate of 95.5% is considered satisfactory. Future research aims to improve the segmentation results and analyze mounting behaviors in greater depth. Additionally, the authors plan to develop decision-making rules for predicting mounting behaviors using a hidden Markov model and apply machine learning methods for time-series data comparison (Noe et al. 2022).

This study addresses the lack of automated detection methods for piglet suckling behavior, which is crucial for assessing piglet live ability and health. The authors propose a two-step computer vision-based detection method for piglet suckling behavior. In the first step, an anchor-free deep learning network is utilized for instance segmentation of individual sows and piglets. Piglets are detected by a localization head, and features are extracted from the region of interest (ROI). A novel attention graph convolution-based structure is applied to distill element-wise features from the extracted features. These distilled features are then encoded by a graph convolutional network and input into the boundary head and mask head for piglet contour and mask prediction. The second step involves tracking the piglets adhered using intersection over union (IOU) calculations between adjacent frames. Piglet motion features are derived from the IOU sequence within 21-frame (3-second) independent processing units, including maximum, minimum, variance, and average values. These motion features are input into a support vector machine (SVM) for classifying a piglet as suckling or nonsuckling. The dataset used for training and verification consists of short video clips (1 minute, 7 frames per second) and one 8-hour long video episode, captured from seven pens of Large White sows and piglets. The proposed method achieves favorable detection performance with an F1 score of 93.6%, recalls of 92.1%, and precisions of 95.2% in short video clips, demonstrating the feasibility of detecting piglet suckling behaviors using modal instance segmentation. In the 8-hour video episode, the time budgets for more than half or all the piglets exhibiting suckling behavior are calculated as 65.6 minutes and 1.1 minute, respectively, representing 88.6% and 1.5% of the total suckling time. These results highlight the superiorities of automated piglet suckling behavior detection and analysis based on amodal instance segmentation. In summary, the study presents an automated detection method for individual piglet suckling behaviors using computer vision techniques. The proposed method achieves favorable detection performance and demonstrates the advantages of automated detection and analysis based on modal instance segmentation. (Gan et al 2022)

In this study, a novel framework for human instance segmentation called Pose2Seg is introduced. The standard approach to this task involves object detection followed by segmentation within the bounding boxes. However, little attention has been given to the unique characteristics of the "human" category, which can be accurately represented by the pose skeleton. The proposed framework leverages the human pose skeleton to distinguish instances with heavy occlusion more effectively than using bounding boxes. The authors demonstrate that the pose-based framework outperforms the state-of-the-art detection-based approaches in terms of accuracy for human instance segmentation and its ability to handle occlusion. To address the lack of publicly available datasets with comprehensive annotations for heavily occluded humans, a new benchmark dataset called "Occluded Human (OCHuman)" is introduced. This dataset contains 4731 images with 8110 detailed annotated human instances, including bounding boxes, human pose, and instance masks. With an average 0.67 MaxIoU for each person, OCHuman is considered the most complex and challenging dataset in the field of human instance segmentation, highlighting the importance of studying occlusion as a research problem. The study concludes by outlining the key components of the proposed pose-based framework, including the Affine-Align operation for selecting Region of Interests (RoIs) based on pose instead of bounding boxes. Furthermore, the human pose skeleton feature is explicitly concatenated with the image feature within the network to enhance performance. The results show that the pose-based system outperforms traditional detection-based frameworks in general cases and exhibits better performance in handling occlusion. Overall, this study contributes to advancing the field of human instance segmentation by introducing a pose-based approach and a challenging benchmark dataset for occlusion-related research (Zhang et al., 2019).

In this study, the researchers proposed an innovative box-free bottom-up approach to address the challenges of pose estimation and instance segmentation of individuals in images containing multiple people. It introduced the PersonLab model, which is a single-shot model that efficiently handles both semantic-level reasoning and object-part associations through part-based modeling. The model utilizes a convolutional network that learns to detect individual keypoints and predict their relative displacements, enabling the grouping of keypoints into person pose instances. Additionally, the authors propose a part-induced geometric embedding descriptor that associates semantic person pixels with their corresponding person instance, thereby delivering instance-level person segmentations. Notably, their system is based on a fully convolutional architecture, facilitating efficient inference regardless of the number of people present in the scene. By training their system solely on COCO data, the authors achieve impressive results. They report a key point average precision of 0.665 using single-scale inference and 0.687 using multi-scale inference on the COCO test-dev dataset, surpassing the performance of all previous bottom-up pose estimation systems. Moreover, their approach demonstrates competitive results in the person category for the COCO instance segmentation task, achieving a person category average precision of 0.417. This study marks a significant advancement in person pose estimation and instance segmentation, offering valuable insights for related projects in the field (Tompson et al., 2018).

The study introduces a model called MaskLab, which addresses the problem of instance segmentation by combining object detection and semantic segmentation. The model generates three outputs: box detection, semantic segmentation, and direction prediction. By utilizing the Faster-RCNN object detector as a foundation, MaskLab accurately localizes object instances through predicted boxes. Within each region of interest, the model performs foreground/background segmentation by integrating semantic segmentation and direction prediction. Semantic segmentation helps distinguish between different semantic classes, including the background, while the direction prediction estimates the direction of each pixel towards its corresponding center, enabling the separation of instances belonging to the same semantic class. Additionally, the authors incorporated recent successful techniques from both segmentation and detection, such as atrous convolution and hypercolumn. The proposed MaskLab model is evaluated using the COCO instance segmentation benchmark and demonstrates comparable performance to other state-of-the-art models. The study presents MaskLab as a promising approach to solving the instance segmentation problem and provides valuable insights into its effectiveness through rigorous evaluation on challenging datasets (Herman et al., 2018).

The study introduces a novel approach called Point-Set Anchors for tasks like object detection, instance segmentation, and human pose estimation. It challenges the conventional method of regressing bounding boxes or keypoints from a central point on the object or person, arguing that this approach lacks comprehensive information due to factors like object deformation and scale/orientation variation. Instead, the proposed method suggests regressing from a set of strategically positioned points that provide more informative features for accurate prediction. The study also incorporates the anchor box technique to generate diverse point-set candidates, considering transformations such as scale, aspect ratio, and rotation. The authors evaluate the proposed framework, Point-Set Anchors, across different tasks and demonstrate its competitive performance compared to state-of-the-art methods. In conclusion, the study presents Point-Set Anchors as a generalized and extended version of classical anchors, showcasing their effectiveness in high-level recognition tasks like instance segmentation and pose estimation. Additionally, the authors introduce PointSetNet, which integrates the proposed point-set anchors into the RetinaNet model and incorporates a parallel branch for keypoint regression, further demonstrating the versatility of the point-set anchor approach through competitive experimental results in object detection, instance segmentation, and human pose estimation (Wei et al., 2020).

This study presents a real-time instance segmentation model called YOLACT. The model achieves a mean average precision (mAP) of 29.8 on the MS COCO dataset while maintaining a high frame rate of 33.5 frames per second on a single Titan Xp GPU. This performance surpasses previous approaches in terms of speed. Remarkably, the model achieves these results using only one GPU for training. The YOLACT model breaks down the instance segmentation task into two parallel subtasks. The first subtask involves generating a set of prototype masks, while the second subtask focuses on predicting per-instance mask coefficients. By linearly combining the prototypes with the mask coefficients, the model produces accurate instance masks. Notably, this process does not rely on repooling, which contributes to the generation of high-quality masks and ensures temporal stability. An analysis of the prototypes reveals that they possess the ability to independently localize instances in a translation variant manner, despite being fully convolutional. This finding highlights the model's capability to learn and represent spatial information effectively. Additionally, the researchers introduce Fast NMS, a more efficient alternative to standard NMS (Non-Maximum Suppression) that reduces processing time by 12 milliseconds, with only a slight decrease in performance. Overall, the YOLACT model demonstrates significant advancements in real-time instance segmentation, achieving impressive results in terms of both accuracy and speed, making it a valuable contribution to the field (Zhou et al., 2019).

This paper presents a novel approach called deep snake for real-time instance segmentation using a contour-based technique. Instead of directly predicting object boundary point coordinates from an image, deep snake utilizes a neural network to iteratively deform an initial contour to match the object boundary, employing a learning-based approach inspired by snake algorithms. The method introduces circular convolution to exploit the cycle-graph structure of a contour for improved feature learning compared to generic graph convolution. The authors propose a two-stage pipeline for instance segmentation, consisting of initial contour proposal and contour deformation stages, which can effectively handle object localization errors. The experimental results demonstrate that the proposed approach achieves competitive performance on various datasets, such as Cityscapes, KINS, SBD, and COCO, while maintaining real-time efficiency with a speed of 32.3 frames per second on a 1080Ti GPU for 512×512 images. In conclusion, this study introduces a learning-based snake algorithm for real-time instance segmentation, leveraging circular convolution for efficient contour feature learning and regressing vertex-wise offsets for contour deformation. The developed two-stage pipeline exhibits superior performance compared to direct regression of object boundary coordinates. Additionally, the authors propose a multi-component detection strategy to overcome the limitation of the contour representation in outlining only one connected component, showcasing its effectiveness on the Cityscapes dataset. The proposed model achieves competitive results on various benchmark datasets, demonstrating its potential for real-time instance segmentation applications (Jiang et al., 2020).

In this study, the researchers introduce a novel method called PolarMask for single-shot instance segmentation. The approach is anchor-box free, fully convolutional, and can be easily integrated into existing detection methods. The main idea behind PolarMask is to represent instance segmentation as the prediction of an instance's contour through instance center classification and dense distance regression in a polar coordinate system. The authors propose two effective strategies to address the challenges of sampling high-quality center examples and optimizing dense distance regression, resulting in improved performance and simplified training. Remarkably, PolarMask achieves a mask mean Average Precision (mAP) of 32.9% on the demanding COCO dataset using a single model and single-scale training/testing, without the need for complex modifications. The study demonstrates that instance segmentation can be approached with the same simplicity and flexibility as bounding box object detection while achieving competitive accuracy. The researchers hope that the PolarMask framework will serve as a strong baseline for single-shot instance segmentation tasks, providing a solid foundation for further research in the field. Generally, PolarMask presents a straightforward and efficient approach to anchor-box instance segmentation, offering valuable insights for related projects (Xuebo et al., 2020).

This study addresses the challenge of construction workers failing to recognize a significant portion of hazards in construction environments. To assist workers and safety managers in identifying hazards, the study proposes a framework for an automated system that detects hazardous conditions and objects in real-time.The framework comprises three independent pipelines: worker localization, semantic segmentation of the visual scene, and detection of static and dynamic hazards. It aims to augment the hazard detection ability of workers and safety managers, offering improved real-time worker localization and an efficient architecture for integrating pipelines.A proof-of-concept system based on the proposed framework was developed and tested in indoor and outdoor construction environments. The system achieved over 93% accuracy in hazard detection. It utilizes live video captured through wearable cameras to track worker positions and provide real-time warnings if they are in proximity to static or dynamic hazards.The system not only assists with hazard detection but also generates large-scale data capturing workers' safety behaviors and near-miss incidents. It addresses the limitations of existing SLAM (Simultaneous Localization and Mapping) methods for localization and proposes a transfer learning approach for retraining construction image datasets with limited data.Furthermore, the system has the potential to analyze eye-tracking data of workers in real-time and can be extended to automate and scale up personalized training and monitoring approaches.In summary, this study presents a framework for an automated system that detects hazards in real-time in construction environments. The system achieves high accuracy and offers computing contributions in worker localization, hazard detection, and data generation. It provides opportunities for personalized training, monitoring, and analysis of safety behaviors in construction workplaces.(Jeelani et al. 2021)

This research focuses on addressing safety issues related to tower cranes on construction sites, particularly accidents caused by falling heavy objects and workers being struck by such objects. The study records video data using a tower crane camera, labels the images, and applies image recognition using the MASK R-CNN method. The identified mask layer is further processed for RGB color extraction to obtain pixel coordinates of workers and dangerous zones. These coordinates are then converted into actual distance measurements to determine safety distances. The research contributes to safety in two ways: first, by establishing an automatic collection, analysis, and early-warning system that does not interfere with workers' normal activities, and second, by improving the safety operation of tower crane drivers through the proposed automatic inspection system. Previous research has focused on protecting workers from hazard zones; however, practical application has been limited due to price and accuracy issues. In this paper, the authors utilize on-site cameras and a Mask R-CNN-based image recognition program to identify field workers and hazard sources without the need for additional hardware. A color layer is created on the identified objects, facilitating subsequent analysis. The method achieves high recognition accuracy and reliability, as indicated by the extracted loss value and average precision (AP) value. By adjusting different RGB thresholds, the pixel coordinates of the mask layer are extracted, allowing for the determination of the maximum pixel distance between hazard sources and the minimum distance between workers and hazard sources through a cross-iterative method. The study successfully develops a high-precision safety distance evaluation model, which can provide assistance to tower crane drivers in maintaining safety (Yang et. al , 2019).

The welfare of sheep during live exports has become a subject of public concern, leading to extensive research focused on monitoring and improving animal welfare. Stocking density is a crucial factor affecting sheep welfare, and its impact can be assessed through the analysis of sheep behavior, position, group dynamics, and physiology. In this paper, the authors demonstrate the application of the instance segmentation method Mask R-CNN to support the recognition of sheep behavior, particularly standing and lying, in different group sizes and over time. The method achieves a validation set mean Average Precision (mAP) of 94% or higher, indicating its effectiveness in identifying sheep behaviors. Further data analysis will provide insights into space requirements for additional sheep allocation and enable daily behavior monitoring to detect abnormal cases, ultimately aiming to enhance the health and well-being of sheep during transportation. Sheep behavior analysis is not only crucial for monitoring and improving sheep welfare during live exports but also in other situations where sheep are confined in limited spaces. To evaluate the effectiveness of computer vision technologies in sheep behavior analysis, the authors focused on the detection and segmentation of standing and lying behaviors using regular RGB 2D cameras. The Mask R-CNN model's performance was assessed on three manually labeled datasets captured by different cameras on different days. The achieved results show a sheep detection and segmentation mAP of 97% or higher, with a slight decrease in performance when additional behavior classification is introduced (94.6%). The accuracy of the model's performance using CCTV images is similar to that achieved with depth cameras, demonstrating promising results for large-scale sheep behavior analysis. In future work, the authors aim to extend this model to differentiate more livestock behaviors, such as drinking, feeding, grooming, and locomotion (Xu et al. 2021).

Ship accidents often occur due to the negligence and inattention of navigators on duty. To address this issue, this paper proposes a human behavior recognition method based on Mask R-CNN and a knowledge-based model using CCTV records from the ship's bridge. The method begins by using camera images as input and extracting keypoints through a convolutional neural network. The feature distribution of these keypoints is then analyzed using clustering algorithms, enabling real-time online recognition of common behaviors such as walking, standing, sitting, steering, looking through a mobile phone, looking out, and napping. The proposed method is validated on samples from a ferry ship in Zhoushan, Zhejiang, China. The validation demonstrates a one-time recognition accuracy of 82% and a rolling recognition rate of almost 99%, indicating significant improvements in navigation safety. The intelligent detecting approach not only identifies high-risk behaviors of ship officers but also triggers early warnings, thereby enhancing the level of navigation safety (Chen & Da, 2022).

The paper introduces an automatic computer-vision approach for detecting individuals traversing structural supports during construction activities to address safety concerns. The approach utilizes a Mask R-CNN algorithm to identify people and recognize their relationship with concrete/steel supports. The results demonstrate the accuracy of the Mask R-CNN in detecting unsafe behavior. The authors propose that this computer-vision approach can be used by site management to identify and provide feedback on unsafe behavior, allowing for immediate intervention and behavior modification. The paper suggests further research to optimize the algorithm for different backgrounds and improve accuracy by reconstructing a three-dimensional construction model (Fang et. al, 2019).

This study focuses on addressing falls from height and scaffolding-related accidents in the construction industry by proposing a correlation-based approach for mobile scaffold safety monitoring and detecting worker's unsafe behaviors. The approach utilizes a deep neural network, Mask R-CNN, combined with an object correlation detection (OCD) module to classify and segment worker's tasks and identify unsafe behaviors. The evaluation of the approach on real scenarios yielded an overall accuracy of 86%, with precision and recall rates of 85% and 97% for safe behavior (class-1) and 91% and 65% for unsafe behavior (class-2). The study highlights the importance of vision intelligence applications in enhancing safety management in construction but acknowledges the existence of limitations that should be addressed in future research (Khan et. al, 2021)

This paper focuses on the detection and recognition of human falling actions using a modular pipeline for generating synthetic data of digital human interactions with a 3D environment. The research demonstrates the use of procedural generation and physics modeling to create a synthetic dataset with realistic movements and falls. The pipeline utilizes Unreal Engine for simulation and captures rgb and segmentation rendering maps, including hit coordinate masks, to represent human-object interactions. The generated data is used to train the Mask R-CNN framework, achieving a high accuracy of 97.6% in recognizing fallen individuals and classifying the type of impact, such as hitting the head. The proposed method addresses the challenge of collecting large-scale annotated datasets for human interaction recognition while avoiding manual annotation errors. However, the study acknowledges the need to increase the number of simulations and variance parameters to improve the prediction of hit map locations in future research (Zherdeva et. al, 2021)

The study presents an approach to address the challenges in social human-robot interaction by integrating deep neural networks with a mechanical robotic system. The proposed method combines the capabilities of Mask R-CNN, a neural network for object detection, with a parallel mini-manipulator mechanism to enable robust interaction between humans and robots. The Mask R-CNN network effectively localizes human faces, which can be used to guide the movements of the robot's head. The integration of the deep learning module into the human-robot interaction system has shown promising results, even when trained with a limited number of samples. The accuracy of the detection and segmentation tasks has improved, making the approach suitable for tasks such as robotics grasping and manipulation, where precise segmentation is crucial. Despite the limitations of the study, such as inadequate training samples, the findings highlight the significant improvements that neural networks can bring to the field of robotics vision, surpassing traditional feature-based approaches. The study also acknowledges the challenges of deploying these advanced tools on mobile platforms and emphasizes the need for further research in this area. The proposed approach holds potential for enhancing human-robot interaction and has implications for future work in robotics and on the computer vision (Huynh et al., 2018).

The paper proposes a dynamic video classification system for fall detection using a single surveillance camera, specifically targeting the elder population. The framework consists of two steps: (1) extracting the human body silhouette from video frames using Mask R-CNN, and (2) utilizing a combination of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) to capture long-term dependencies between successive frames. The system achieved impressive results, with a model using 10 frames per video achieving 100% accuracy, precision, recall, and F1-score for fall detection. The proposed system offers a non-intrusive solution for fall detection in elders by utilizing a single camera and avoiding complex devices. The use of instance segmentation with Mask R-CNN and LSTM enables effective learning of dependencies between frames. The system's performance demonstrates its potential to accurately detect falls while minimizing the impact on the lifestyle of the elders. Furthermore, the paper suggests future developments, including scaling the system to manage multiple buildings for regional-level supervision of elder activities. It also highlights the potential for using the system to address other problems and enhance telemedicine for indoor elder care (Mobsite et al., 2020).

2.3 Conceptual Framework

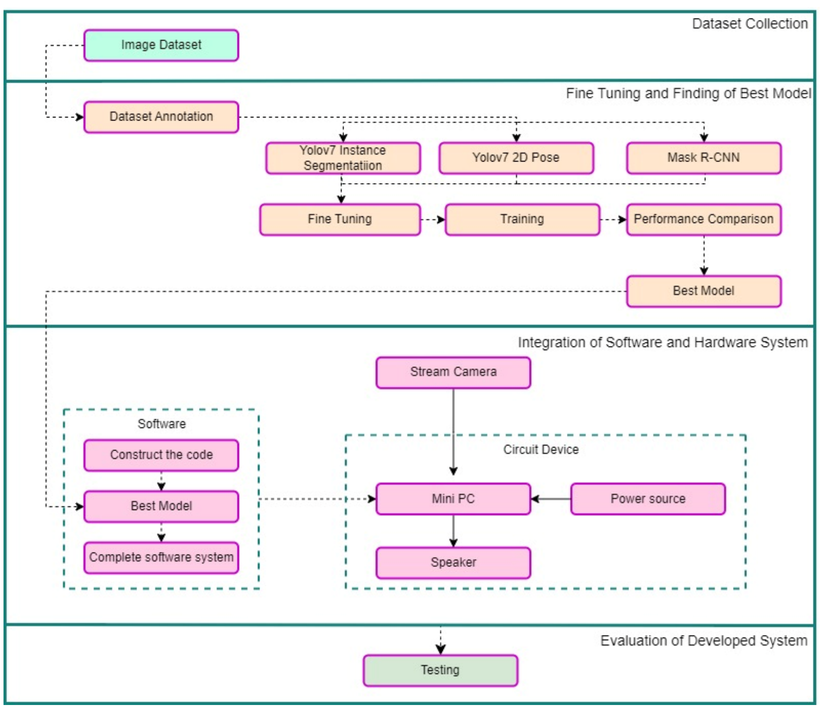


Figure 1: Conceptual Paradigm

The framework in figure 1 shows the phases in the development of the project. The first phase is the image dataset collection. Next is fine tuning and finding the best model where the researcher would fine tune the pretrained model to meet the specific function for detection. After this, it is compared to different fine-tuned models based on performance. This phase includes training and testing of different models to be able to achieve the objective. Once the best model is chosen among the models, integration of software and hardware systems takes place. For the software, it includes the process of constructing the software where the model is integrated. Next is the hardware system. The input is a stream camera that has real-time image acquisition. It is extracted from the mini-PC being controlled by the software. The power source kept the main driver and maintained its life. The last phase is the evaluation where the developed system was tested. When the system detected possible accidents based on safety hazards, the system triggered the alarm. The teacher is provided with an access live view video of what is happening around the students.

Furthermore, the system model detects the possible accidents the kids might face based on the safety hazards that the kid interacted with, then alarm the teacher once the system detects a possible accident that might happen. These include the following:

A. Falls due to climbing on the:

   a) Chairs

   b) Tables

B. Trips due to sprinting or running.

C. Electrocuted due to tampering of outlet.

CHAPTER III

DESIGN AND METHODS

Design and Methods

This chapter presents the constraints, experimental design, and procedure, equipment/facilities/program, data collection and treatment, budget requirements, expected output and the Gantt chart respectively.

3.1 Multiple Design Constraints and Trade-Offs

In this section, the different design constraints in terms of global, environmental, societal, economic, safety and manufacturability are presented. These constraints are important as this sets the conditions required to complete the design project. Through this understanding, it was able to determine the limiting factors that are important in the development of the design project and how it was addressed those limitations. It also helps to narrow down choices in the development of the design project. These designs include alternatives that were used in the development of this design project and the evaluation of which design to use was based on different constraints that these designs have.

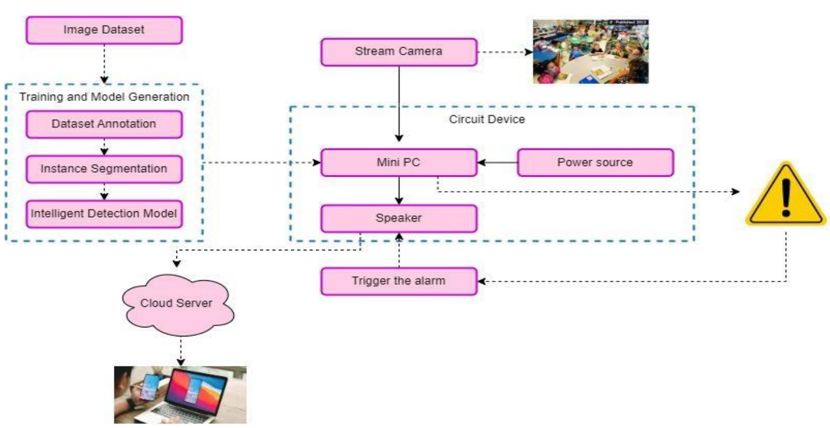


Figure 2: Design 1

Design 1 in figure 2 shows the whole framework of the system with a different approach in terms of output and algorithm compared to the following design. The framework shows the input which comes from the stream Camera. The frame was extracted from this through a mini-PC which was controlled by software that was analyzed by the detection model trained by the image datasets. Once the intelligent mechanism detects safety hazards present, it makes use of the speaker which was triggered to alarm the teacher. The power source serves as the main driver of the circuit and the teacher was also be able to access the live stream video whenever these are out of the classroom.

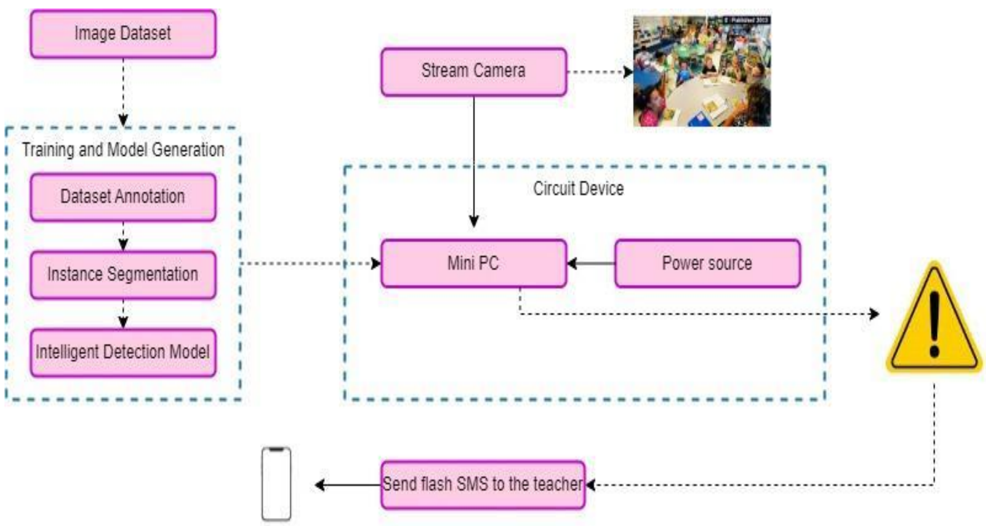


Figure 3: Design 2

Design 2 presented in figure 3 was similar when it comes to input, but the algorithm and the output are different. If Design 1 simulates the output through an alarm, Design 2 simulates the output through flash emergency SMS which alerts the teacher once the intelligent system detects the possible accidents depending on safety hazards. This implies that the system doesn't need an alarm device, instead, it only requires a smartphone for the teacher.

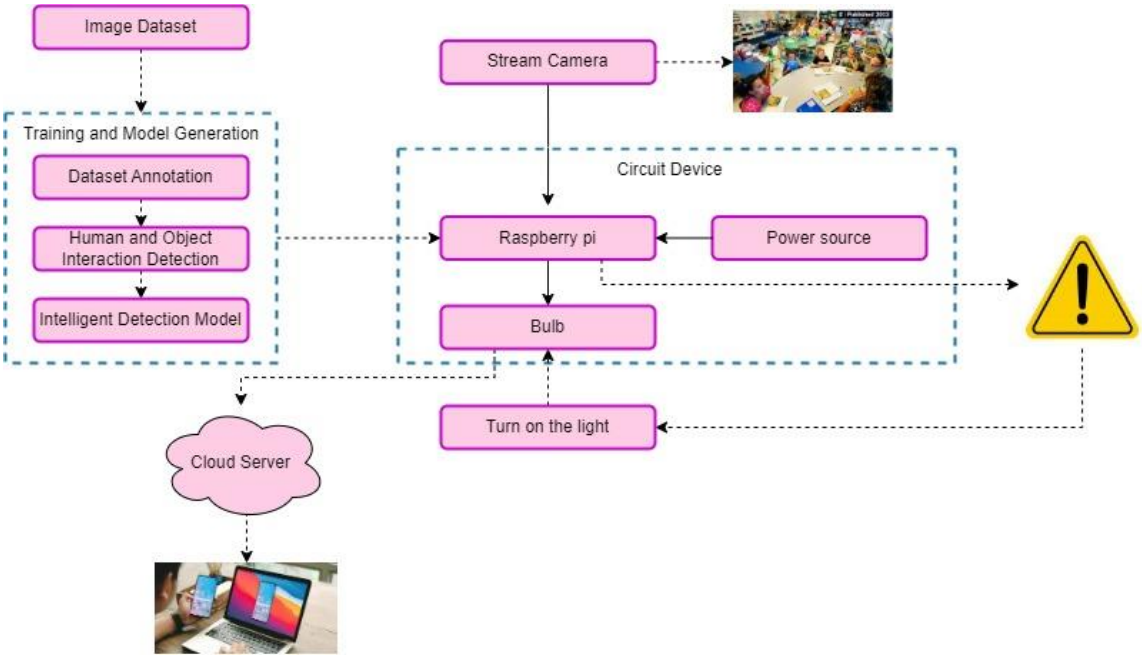


Figure 4: Design 3

Basically, design 3 was like design 1 but with different materials and output. Shown in figure 4 was the output design which consists of a light that was turned on and a smartphone for the teacher to access the live video stream of the kids in the classroom. Also, raspberry pi was used for this design instead of the mini-PC. To further explain, once the system detects that there was a possible accident present, the system produces an output light. Further, if the teacher chooses to view the live video of the kids wherever outside the classroom, it ables to view it through the system.

$$\text{Rate} = \frac{maximum - actual}{maximum - minimum} \times 4 + 1$$

To determine which type of design that was used, equation 1 was utilized to rate the different constraints and determine the one that ranks the highest. This was based on the chosen rating which was 1-5 where 5 ranks the highest and 1 was the lowest. This formula includes the maximum value of the constraints then minus the actual value divided by the difference of maximum and minimum value times 4 plus 1.

Table 1: Maximum and Minimum Value of Constraints

Constraints	Minimum Value	Maximum Value
Global	1	5
Economic	10000	16000
Societal	1	5
Environmental	1	5
Safety and Health	1	5
Manufacturability	5 months	8 months

Table 1 presents various constraints with their minimum and maximum values. These constraints include global considerations (1-5), economic factors (10,000-16,000), societal implications (1-5), environmental concerns (1-5), safety and health (1-5), and manufacturability (5-8 months). These ranges indicate the boundaries within which decisions and actions should be made, considering global impact, economic thresholds, societal and environmental consequences, safety and health priorities, and manufacturing timelines. Adhering to these constraints ensures responsible decision-making and sustainable outcomes. This maximum and minimum value was used to determine the best design suitable for the objectives of this study. It was the basis of the chosen design to ensure the achievement of the objectives.

Table 2: Multiple Constraints Decision Matrix

Design		1	2	3
Global		**2**	2	1
10%	**Rate**	**4**	4	5
	Score	**0.4**	0.4	0.5
Economic	**Cost**	**15000**	14000	12000
	Rate	**1.67**	2.3	3.67
30%	**Score**	**0.50**	0.7	1.1
Environmental		**2**	2	2
	Rate	**4**	4	4
10%	**Score**	**0.4**	0.4	0.4
Societal		**3**	2	1
	Rate	**3**	4	5
10%	**Score**	**0.3**	0.4	0.5
Health and Safety		**4**	5	1
	Rate	**5**	1	2
30%	**Score**	**1.5**	0.3	0.6
Manufacturability	**Speed**	**6 months**	**6 months**	**7 months**
	Rate	**3.67**	3.67	2.33
10%	**Score**	**0.367**	0.367	0.233
Total		**3.467**	2.567	3.333

Global Constraints (10%)

Global constraints in table 2 were set to 10% since the project design mainly focused on economic and environmental safety. This constraint means consistency and consists of a set of constraints. In constructing a design project, one thing that needs to be considered was the consistency of the design which means the result, effect, and impact needs to be consistent. Another thing was the implementation of modern design that would fit the global economy and invention.

Economic Constraints (30%)

The economic constraint in table 2 was the most significant limiting factor in this design project. Economic feasibility was an important aspect that needs to be maintained considering the limited budget for the project which ranges from P10,000-P16,000. In terms of economics, it must find the best cost fitted for the proposed budget and at the same time still be efficient for the project design and its accuracy. To further achieved it, open-source data was utilized instead of data that requires payment.

Environmental Constraints (10%)

Environmental constraints in table 2 need to be considered especially for accurate detection of objects and hands whether it was slightly dark or blurry but since the camera that was used in the three designs was both the same, it yields the same results and does not significantly impact the choice of the best design.

Societal Constraints (10%)

The rate for societal constraints shown in table 2 was 10% since the accessibility of this device was mainly for kids attending kindergarten. It was also limited to the inside of the classroom. Further, the system implemented would be user-friendly and efficient for the school.

Health and Safety Constraints (10%)

Health and safety shown in table 2 are one of the most significant as this project focused on the safety of people specific children attending kindergarten. All the components and devices should be placed properly, especially the stream camera. In terms of the system, it was considered the most efficient and accurate for maintaining the safety of the user. Further, accuracy was also considered in these constraints, considering that the more accurate it was the more it can prevent and provide intervention from danger around kids.

Manufacturability Constraints (10%)

Another constraint shown in table 2 that needs to be considered was manufacturability. Since there was a specific time for the project design, the possible time for it to be done needs to be determined to choose a design that fits the specific time range set and not the design that would exceed the time limit given.

Trade-Offs

The design presented in the previous sections differs from components, algorithms, and most especially the output. The first design makes use of an alarm device which in terms of output was ring correspondingly and prove to have shorter manufacturability compared to others. It utilizes instance segmentation. Next was the implementation of flash SMS which lessens the cost because it does not require any alarm device. The downside was the SMS might not be efficient for it may not effectively alert the teacher considering the teacher focused on teaching. Further, it utilizes the 3D pose estimation of multi-person and object. Lastly, it utilizes the yolov7 algorithm where human and object interaction was the basis of detection.

As the multiple constraints analyzed it was concluded that the best design to be used was Design 1, which implements alarm system output. Design 1 falls into the desired range for costs which are P10000-P16000. Also, it has a significantly high score in terms of health and safety which are also important constraints.

This also satisfies the global, environmental, and manufacturability which was shorter than the other. In this design, the alarm was not just for alerting the guardian, according to Online School Center (2021), “Noise essentially diverts the mental resources of the introvert’s brain that was responsible for memory recall and problem-solving and distracts”. This implies that noise can also be a factor to divert the child’s attention, delaying and distracting before the guardian arrives at the scene. This act, aside from the teacher preventing the kid from further engaging in a hazard was an effective intervention that increases the possibility of preventing the danger before it causes damage. Even though the other two designs have not chosen, these are still significant as these are the basis for further improvement.

3.2 Technical Design and Procedure

Engineering Standards

Table 3. Technical Specifications of Mini-PC

Processor	i5-6500T 6th Gen 2.5GHz
RAM	16GB
Storage	500GB HDD

Continuation:

Operating System	Windows 10
USB Ports	6 (4 rear, 2 front)
VGA Port	1
Display Port	2
Audio Ports	2 (front)
Type C Port	1 (front)
Ethernet (LAN)	Gigabit
Graphics	Intel HD Graphics 530 Dynamic Video Memory Technology

Table 3 presents specifications that are crucial machine learning, software development purposes and final system design. The processor, an i5-6500T 6th Gen with a clock speed of 2.5GHz, was dedicated to faster detection and machine learning tasks. A higher RAM capacity of 16GB contributes to the speed and efficiency of machine learning programs. The six USB ports (four at the rear, two at the front) are essential for connecting peripherals like keyboards and mouse during software development and simulation. Further, the VGA port enables connection to external monitors, aiding multitasking, and collaboration. The presence of audio ports allows for connecting speakers, which serve as output for hazard detection alarms. The Gigabit LAN facilitates high-speed internet connectivity, particularly useful for live streaming through tools specifically TeamViewer. Lastly, the Intel HD Graphics 530 Dynamic Video Memory Technology handles basic graphics tasks, providing visual support for machine learning applications. Together, these specifications enhance the system's performance and contribute to seamless machine learning, software development experience and final output design development.

Table 4. Technical Specifications of Stream Camera

Look Resolution Ratio	1080P/720P
Video Format	AVI
Consumption	230MA/3.7V
Memory Card Type	TF card
Maximum Capacity of Memory Card	128GB
Frame Rate	60 fps

Table 4 presents the specifications of stream camera which was used in the system and connected in the Mini-PC. Look Resolution Ratio specification states that the system supports a resolution of 1080P/720P. This was important for capturing high-quality video footage in the classroom, allowing for clear and detailed monitoring of the surroundings. A higher resolution enables better identification of potential hazards and enhances the accuracy of hazard detection algorithms. For the video format, which was the AVI, it was commonly used video format that provides good compatibility across various devices and software platforms. It ensures that the recorded video footage can be easily accessed, played back, and shared for analysis and review. Consumption specification indicates that the system operates at 230MA/3.7V. This refers to the power consumption of the system during operation. It was important to consider power efficiency to ensure that the system can operate for extended periods without draining the power source rapidly. This allows for continuous monitoring and hazard detection in the classroom without frequent interruptions.

Furthermore, the memory card type which was specified as TF card (also known as a microSD card). This indicates that the system utilizes a TF card for storing recorded video footage and other data. It allows for easy removal and transfer of recorded data for further analysis or archiving. It mainly contributes to the development process of the system, since it also utilizes this for gathering datasets which needs a recorded video for image datasets. The memory maximum amount of storage capacity supported by the system was 128GB. A larger capacity allows for storing a substantial amount of video footage, ensuring that critical moments or hazardous situations are captured and retained for later reference or analysis. Also, a higher frame rate can provide more frequent updates on the position and movement of objects in a video stream. This aids real-time tracking and improves the accuracy of object detection algorithms used in the system.

Overall, these specifications are relevant for a classroom hazard detection system as it provides high-resolution video capabilities, efficient power consumption, and appropriate storage options. Together, it contributes to the effective monitoring, recording, and analysis of potential hazards within the classroom environment.

Table 5. Technical Specifications of Speaker

Audio Jack	Yes
Amplifier Specification	3W
Power Supply	USB Cable

Table 5 shows the product specification for output device of the system which was the speaker. It was connected to the Mini-PC through the audio jack and USB cable. The presence of the audio jack allows for connecting external audio devices to the speaker system. In this case, it was crucial for the system to have an audio output to produce an alarm sound when a hazard was detected. The speaker can serve as an effective and audible means of alerting the teacher to the presence of a potential danger. While the amplifier specification specifies that the speaker system has a 3W amplifier. This indicates the power output of the amplifier, which influences the volume and clarity of the audio produced. A 3W amplifier can provide sufficient power for producing loud and clear audio, ensuring that the alarm sound was easily audible throughout the classroom. The power supply was provided through a USB cable. This means that the speaker system can be conveniently powered by connecting it to a power source, such as a computer, USB charger, or power bank. The USB cable ensures easy accessibility to power, making it convenient for installation and use in the classroom.

Table 6. Technical Specifications of Mobile Phone

Operating System	Android 7.0 or later
Team Viewer Application	Yes
Internet Connectivity	Yes

Table 6 presents the technical specifications of mobile phones to be used for live streaming the activities of kids inside the classroom. It was necessary to have a mobile phone running Android 7.0 or later as the operating system. This ensures compatibility with the required applications and software. The TeamViewer application was essential for establishing a remote connection and streaming the content from the mobile device. Additionally, a reliable internet connection was necessary to transmit the live video stream of the classroom activities to the desired platform or destination. By meeting these specifications, the mobile phone can be effectively utilized for live streaming, enabling remote viewers to observe and monitor the activities of the kids inside the classroom in real-time.

Process Flow

To ensure that the project was completed successfully, it intended to follow a set of guidelines. This worked on an IoT-based alarm and real-time monitoring system for kindergartens. In this study, it used the Process Flow shown in Figure 5.

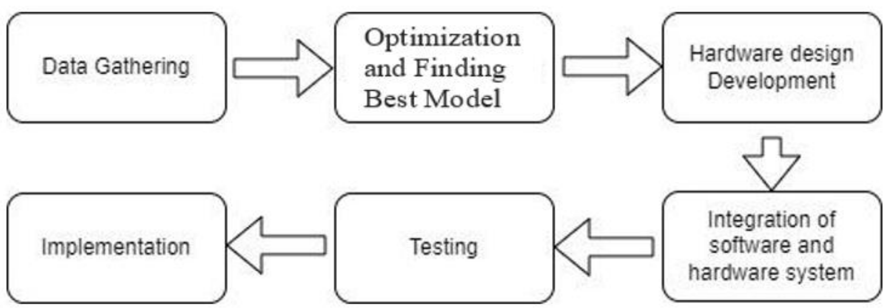


Figure 5. Process Flow

The Data Gathering was the first phase. This phase consists of a review of existing studies and recommendations for how it might improve. This also identifies certain potential issues for which it has developed concrete solutions. Also, it includes dataset gathering for training the model to be integrated into the hardware system.

The second phase of the development of the study was optimization and finding the best model. It began training and testing different models. Throughout this phase, it modified the model until it met the project's design requirements.

The third phase was Hardware Design Development. This began designing a potential hardware prototype during this phase. Throughout this phase, it modified the model until it met the project's design requirements. The design project was a system that consisted of a streaming camera as well as an alarm. The streaming camera was connected to a system that was programmed to detect and analyze situations and if the trained system recognizes a hazard condition, it automatically alerts teachers with a programmable sound.

The fourth phase was integration of software and hardware design. Following the design of the hardware prototype, the software design was integrated in the hardware to achieve the output design.

The fifth phase was Testing. Following the design of the output device, it tested the device and discussed what needs to be done to improve the design when incorrect data was found. This phase was performed frequently with the hardware being modified to match the project's desired data. Similarly, this phase also incorporates planning for the implementation to production.

The study's implementation was the final phase of development. This also finalizes the design and build the device. All final adjustments were made based on the needs standard and requirements.

Equipment/Facilities/Program

This part of the chapter explains and shows all the equipment, facilities, and programs included within the design of the study as well as each of their appropriate uses.

I. Equipment

The equipment used in creating the system includes the laptop or PC which was used significantly for creating, running, testing and execution of the system. Also, the internet was used for the testing of the output system and for researching, collecting and gathering of data needed for the development of the project. Smartphone was also used mainly for testing the live stream view inside the classroom where team viewer app was utilized.

II. Facilities

The proposed system was designed to detect accidents based on the safety hazards present. This does not need a huge facility, a well-ventilated and enough if it would not be a hindrance to the productivity.

III. Programs

The proposed system utilized Python programming language together with the following programs as stated below.

Image Processing. Image Processing was a method to perform some operations on an image, to get an enhanced image, or to extract some useful information from it. It was a type of signal processing in which the input was an image, and the output may be an image or characteristics/features associated with that image.

OpenCV. OpenCV was a great tool for image processing and performing computer vision tasks. An open-source library can be used to perform tasks like face detection, objection tracking, landmark detection, and much more.

Instance Segmentation. It was a computer vision task that enables the precise localization and segmentation of individual human instances in an image, allowing for accurate analysis and tracking of body parts and their movements.

2D Pose Estimation. Estimating the 2D position or spatial placement of key points on the human body from visuals like photos and movies was known as 2D human posture estimation. Different hand-crafted feature extraction approaches are used for each body part in traditional 2D human posture estimate methods.

IV. Model Architectures

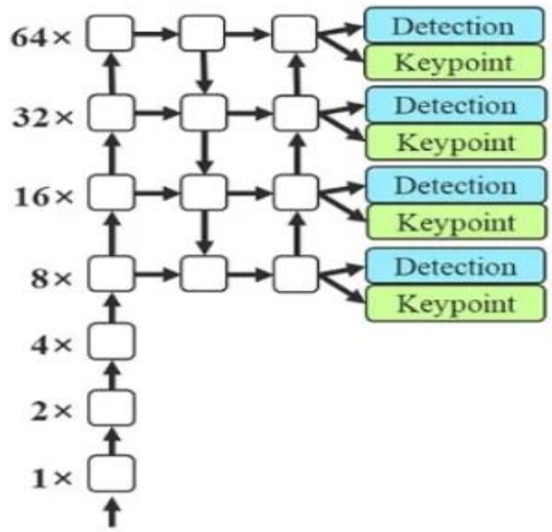


Figure 6. Yolov7 Pose Model Architecture

In contrast to traditional pose estimation techniques, figure 6 shows the YOLOv7 pose which was a single-stage multi-person key point detector. It was comparable to the bottom-up strategy but does not use heatmaps. It was a development of the YOLO-posture one-shot posture detector and combines the best elements of Top-down and Bottom-up methodologies. The COCO dataset, which contains 17 landmark topologies, was used to train the YOLOv7 Pose. It was implemented in PyTorch, which makes it incredibly simple to alter the code to meet the system needs.

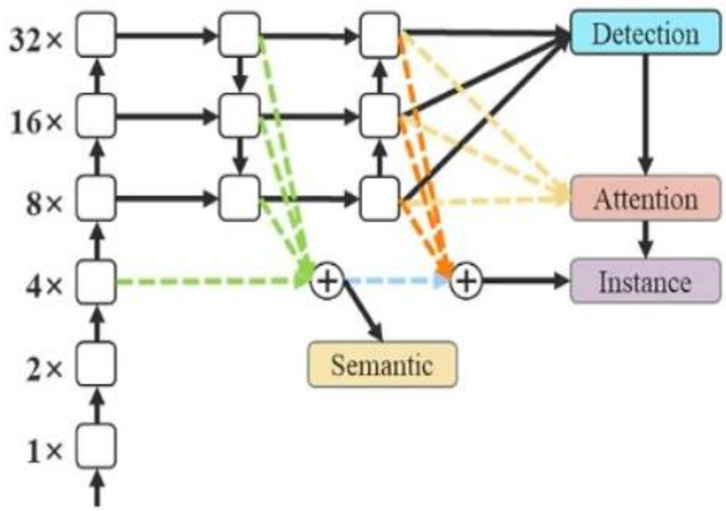


Figure 7. Yolov7 Mask Model Architecture

Figure 7 shows the instance segmentation which was carried out using the YOLOv7 and BlendMask integration. As a result, the MS COCO instance segmentation dataset was used to refine the YOLOv7 object detection model, which was trained across 30 iterations. The technique of segmenting an image involves breaking it up into various areas or pieces that are all members of the same class.

This clustering task was based on standards, like color or texture. This method was additionally known as pixel-level classification. To put it another way, it entails dividing up images (or video frames) into several items or segments.

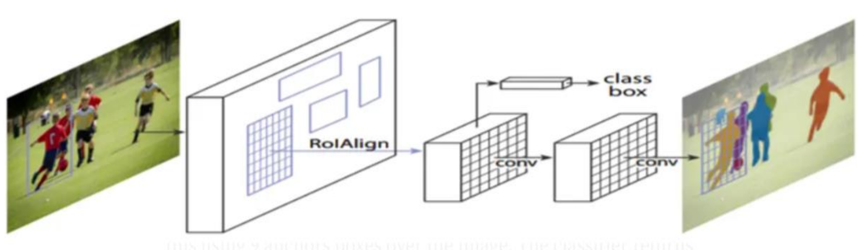


Figure 8. Mask R-CNN Architecture

Figure 8 displays the binary mask classifier which creates masks for each class, and the region proposal network (RPN) suggests possible object bounding boxes. These two components make up the mask R-CNN model. Anchor boxes are used by Mask R-CNN to identify multiple items, objects of various sizes, and objects that overlap in an image. This increases the effectiveness and speed of object detection.

A group of predetermined bounding boxes with a specific height and breadth are known as anchor boxes. These boxes are intended to record the size and aspect ratio of the item classes the team are trying to identify. Mask R-CNN generates hundreds of predictions to predict multiple objects or numerous instances of things in a picture and it was filtered depending on the confidence score.

3.3 Data Collection and Treatment

The data was gathered by compiling ideas from a variety of news articles and peer-reviewed research publications that address the relevance of real-time monitoring of children in preventing dangerous circumstances. Further, the dataset that was used for the model was gathered in a classroom environment to achieve the objective.

The algorithm was used to compare and analyze using various studies and proofs which leads to the instance segmentation being chosen among the algorithms.

3.4 Budget Requirements

The overall cost of the system implemented was presented in Table 7. The budget requirements include the equipment Mini-PC, which was the main controller of the system, then the stream camera which was the input device and finally the speaker which was the output device of the system. The total cost of the gadgets was 15, 480.00 pesos.

Table 7: Budget Requirements

Component	Quantity	Unit Price (PHP)	Subtotal (PHP)
Mini PC	1	7,200.00	7,200.00
Stream Camera	1	8000.00	8000.00
Speaker	1	280.00	280.00
	TOTAL		**15, 480.00**

Shown in table 7 the breakdown of the budget requirements and the total value. This would use mini-PC as the main tool that would act as the brain for the system to function. Most of the components were dependent on this microcontroller because it has the capability to control things like the stream Camera and USB Speaker. The stream camera captures the surroundings in real-time monitoring so the system would detect if the kids were prone to safety hazards. Further, the USB-connected speaker act as an alarm to give awareness to the guardians of the user, it is an essential component of our device because it can alert and inform the teachers as soon as possible; the power supply would trigger the whole system to function. Furthermore, the guardian's smartphone would have access to the live video streaming of kids.

CHAPTER IV

PRODUCT DEVELOPMENT

This chapter provides important information regarding the product's specifications, the development process, and the simulation and testing performed throughout the design project.

4.1 Product Specifications

Table 8. Product Components

PRODUCT COMPONENTS	
 **LAPTOP**	The laptop serves as the central control unit and runs the necessary software for data processing, analysis, and user interface.
 **MOBILE PHONE**	The mobile phone enables teachers to remotely monitor the classroom, receive system alerts, and promptly respond to safety hazards, regardless of their physical presence.
 **STREAM CAMERA**	The streaming camera captures live video of the kindergarten classroom, enabling real-time analysis and monitoring of the environment by the intelligent detection model.
 **MINI PC**	The mini-PC in an IoT-based system functions as an edge computing device, processing and analyzing data locally to reduce the burden on the central control unit
 SPEAKER	The speaker component produces alarm sounds to alert the teacher when safety hazards are detected.

Table 8 presents the product components used for the development of the system. It includes a laptop which was used to develop and simulate software code to be embedded in a hardware system. Also, the mobile phone was included for the testing of the live stream view of the video. Next was the stream camera which was one of the most essential parts of hardware devices. This was mainly used for real-time viewing of what was inside the classroom. Mini-PC was for the main control and served as the computing device which has the embedded software developed. Lastly, speaker was used as an output device which was connected to the Mini-PC to serve as the alarm when the system detected a hazard.

4.2 Product Development

To be able to achieve the given objectives, the development process was divided into 4 phases which includes the dataset collection, determining the best model, integration of software and hardware systems and then testing.

4.2.1 Dataset Collection

Figure 9. Sample Dataset from Classroom Environment

For the first phase which was the dataset collection, shown in figure 9 are sample datasets gathered from the classroom environment in Manalupang San Vicente. Here, the gathered data sets include the different classes including Running, Normal, Climbing and tampering.

4.2.2 Developing and testing of different models.

Upon completion of dataset collection, this proceeds to the next phase which includes the dataset annotation, different model development and comparison.

4.2.2.1 Dataset Annotation and Preparation

Figure 10. Sample Annotated Dataset for Mask R-CNN and Yolov7 Instance Segmentation Models

Figure 10 shows the sample annotation of datasets gathered from the classroom environment. As seen in the picture, a polygon method was used to trace the body of each of the kids in the classroom to train the instance segmentation model. The datasets have been organized using Roboflow which automatically divides the dataset into testing, training, and validation.

Here the dataset was divided into 80% training set, 10% validation set, and 10% testing set. Further steps such as dataset augmentation have been conducted to help with the optimization of the model developed and trained for intelligent detection.

Figure 11. Sample Annotated Dataset for Training Yolov7 2D Pose.

Figure 11 shows the sample annotation for the 2D Pose Estimation. Unlike the two models above, 2D pose focuses only on the key points of the person which was extracted in a CSV file that serves as the dataset for the pre-trained 2D Pose model. In here, the dataset was also divided into 80% training set, 10% validation set, and 10% testing set. Further steps such as dataset augmentation have also been conducted to help with the optimization of the model developed and trained for intelligent detection.

4.2.2.2 Fine Tuning and Comparison of Fine Tuned Pretrained Models

For the model fine tuning, researchers used Google Collaboratory for training and testing the three models specifically Mask R-CNN, yolov7 instance segmentation and 2D Pose.

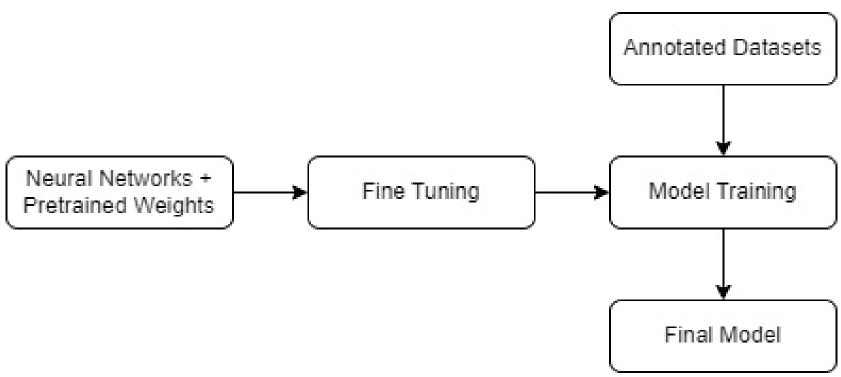


Figure 12. Process Flow for Fine Tuning the Three Models

Shown in Figure 12 the implementation of fine tuning to meet the desired criteria for detection and classification. As shown above, it used pretrained weights which was already trained network for starting point of the training of the new network which leads to fine-tuned model. Once fine tuning was done, it then proceeds to training where in the dataset used was the customized datasets which were annotated based on the previous step.

Table 9. Performance Evaluation of the Three Model

Models	Performance Accuracy	False Positive Rate
Mask R-CNN	94%	12%
Yolov7 Instance Segmentation	97%	9%
2D Pose Estimation	89%	18%

Table 9 presents a performance result comparison of different models that were developed and trained using a dataset that was divided into three parts: 80% for training, 10% for testing, and 10% for validation. The purpose of this comparison was to evaluate and assess the effectiveness of these models in classifying safety hazards. This evident proves that the Yolov7 Instance Segmentation model outperformed the other models in terms of accuracy and false positive rate. This indicates that the Yolov7 model demonstrated a higher level of efficiency and effectiveness in accurately identifying and classifying safety hazards. Notably, it excelled in recognizing hazards such as climbing, running, and tampering.

Based on these favorable results, the decision was made to integrate the Yolov7 Instance Segmentation model into the hardware system. This choice was driven by the model's superior performance and its ability to provide accurate and reliable hazard classification. Furthermore, incorporating this model into the hardware system, it was anticipated that safety measures and precautions can be enhanced, as potential hazards can be identified and addressed more efficiently and effectively.

4.2.3 Integration of Software and Hardware System

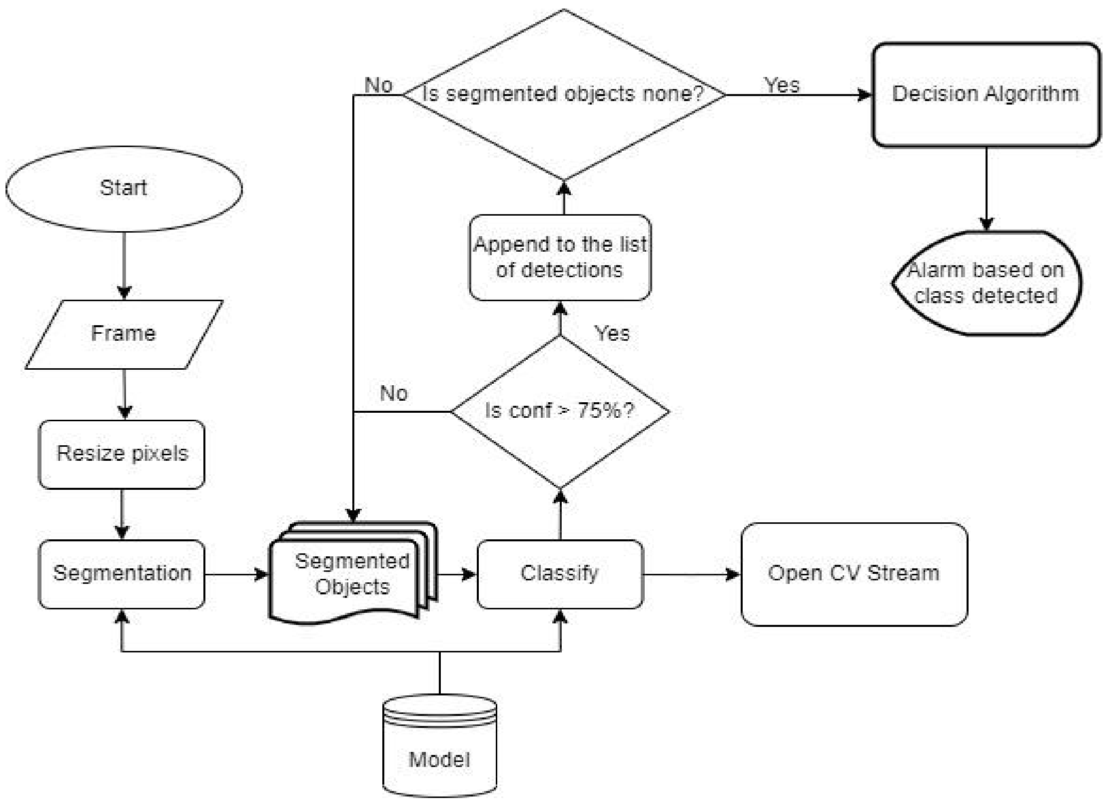


Figure 13. Flowchart of the Software to be Integrated in Hardware System

Figure 13 above shows the complete flow of the code embedded in the hardware device. Once the mini pc was started, the code automatically simulates and does its thing. Now to break down its process, it was designed to first extract from image frame by frame. Once a frame was extracted, the code resizes the pixel of the frame to enable the model to inference faster and avoid the delay.

Furthermore, it undergoes segmentation of objects in the frame which specifically the person the frame. Now the batch segmented objects were further classified using the trained model and once the confidence score of this was higher than 75%, the classification was included otherwise it was disregarded. This was because there was false positive classification in the system that ranged below 75% confidence score and to avoid that researchers have set a specific limit on what was considered as part of the detection.

Once all the segmented objects have been classified in a frame, it performs an analysis of what alarm to be considered depending on the given class. For each class, including the Running, Climbing and Tampering there are different alarms dedicated to each of the classes.

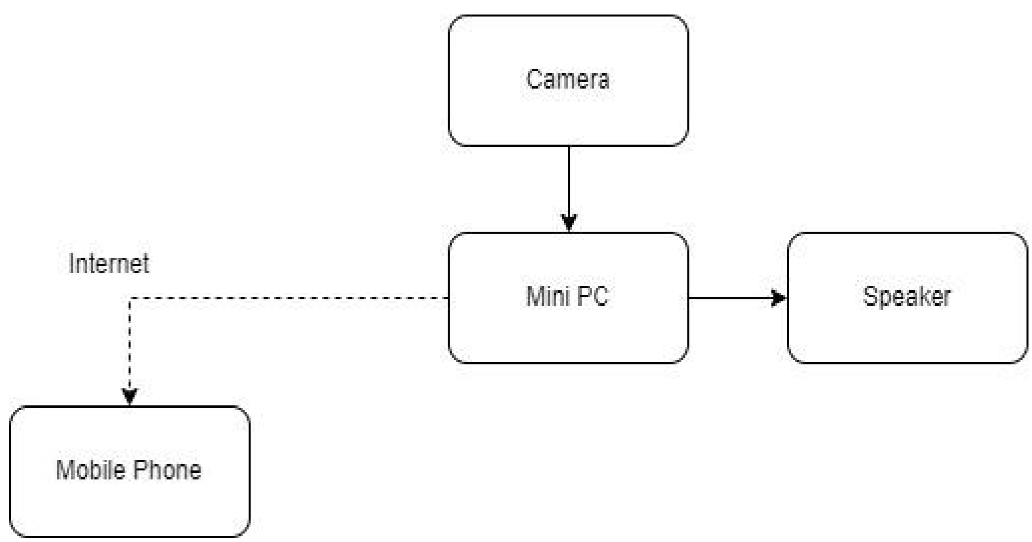


Figure 14. Hardware System Architecture

Figure 14 shows the hardware system that consisted of the Mini PC, speaker and mobile phone which was connected via the Internet. In this system, the camera serves as the input while mobile phone and speaker serves as the output for the system.

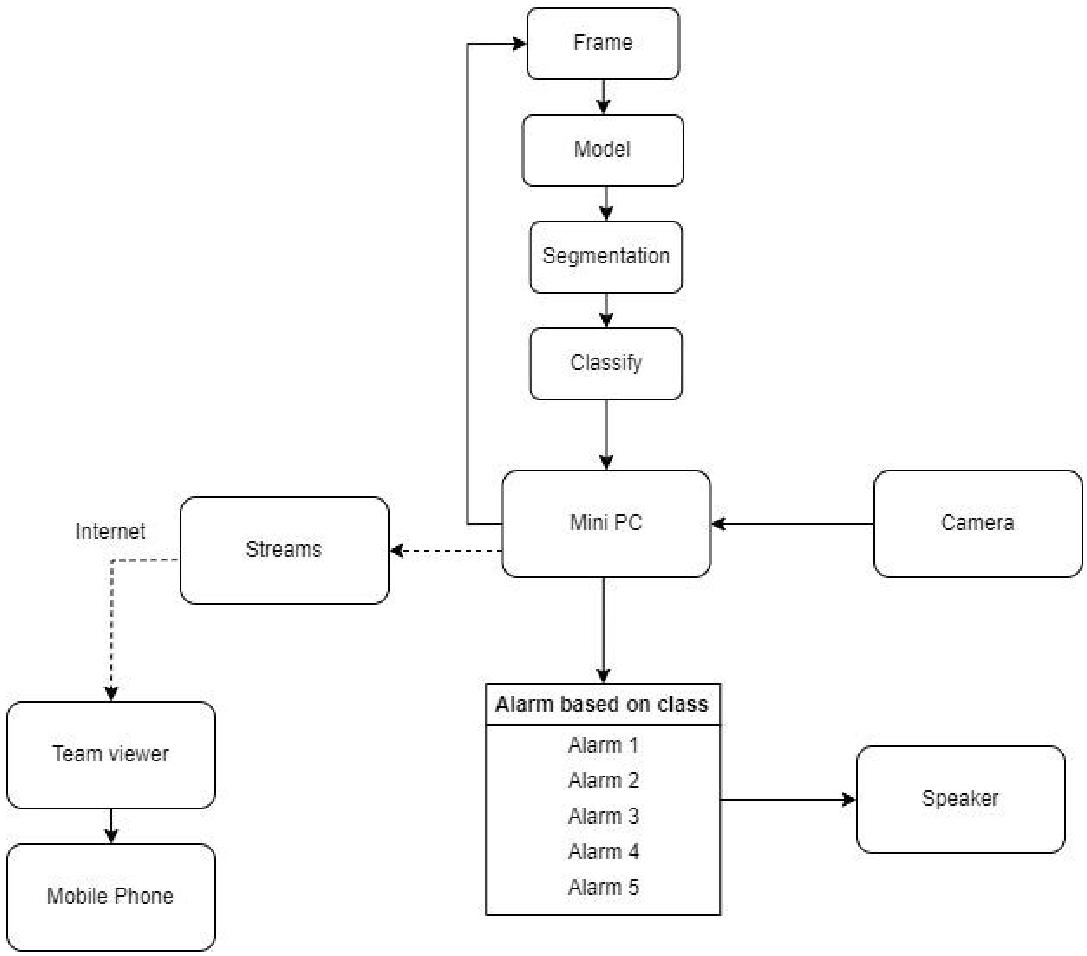


Figure 15. Block diagram of the Integrated Software and Hardware System

Figure 15 shows the system includes Camera as input then was passed to the integrated system which comprises the mini pc and software code. Here, the input was analyzed based on the segmented output and picked the alarm that was suitable for the specific class which was detected by the system. Once the system knows what alarm was suitable it triggers the speaker. Now, for the live stream with the use of a team viewer the teacher can view the live footage of the students in case there was too far in the classroom. To do that, the user first has to login and connect to the computer and once connected it can obtained access the live video of the kids and at the same time it alarmed even in distance.

Figure 16. Output Device with the Integrated Software

Figure 16 displays the final output device which was attached to the upper front center of the classroom in Manalupang San Vicente Elementary school. The output device consists of the speaker which was placed in the right-hand corner together with the mini-PC while the camera was connected to the mini-PC and attached in the center of the front classroom.

4.3 Simulation and Testing

Upon having the best model, specifically Yolov7 Instance segmentation with high accuracy integrated it to their hardware system. Then, it proceed to do a live testing with their model and device in Manalupang San Vicente Elementary School.

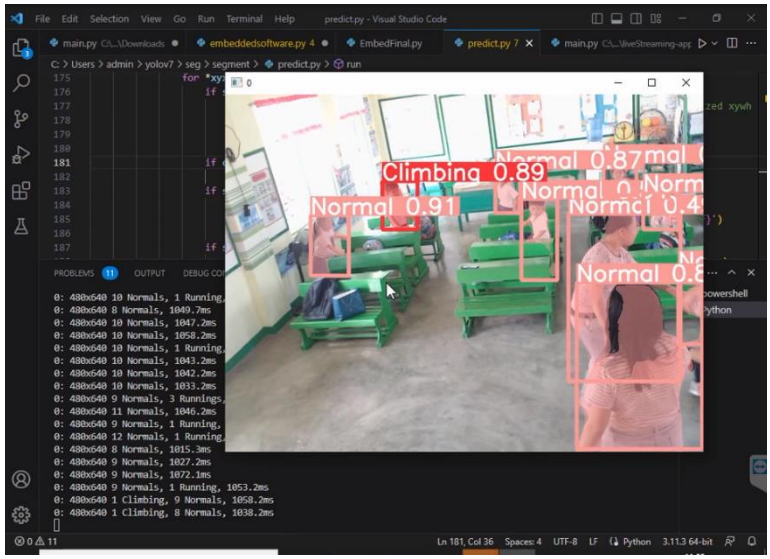


Figure 17. Detection of Climbing and Normal

Figure 17 shows the detection of climbing class upon testing in the classroom environment of Manalupang San Vicente Elementary School. This presents the live stream of kids which undergo instance segmentation and classification with confidence score of 89% and Normal classification that has a confidence score that plays around 87% to 94%.

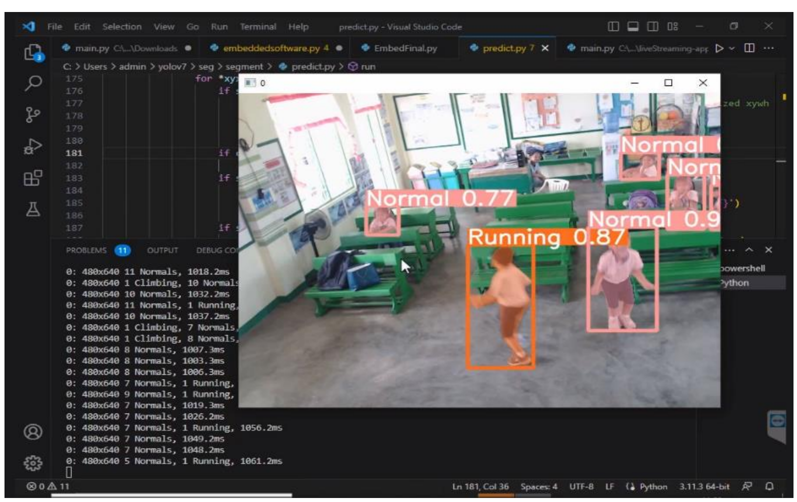


Figure 18. Detection of Running and Normal and Normal

Figure 18 shows the detection of running class upon testing in the classroom environment of Manalupang San Vicente Elementary School. This displays the live stream of kids which undergo instance segmentation and classification of running with an accuracy of 87% and normal classification that plays around 77% to 90% or above.

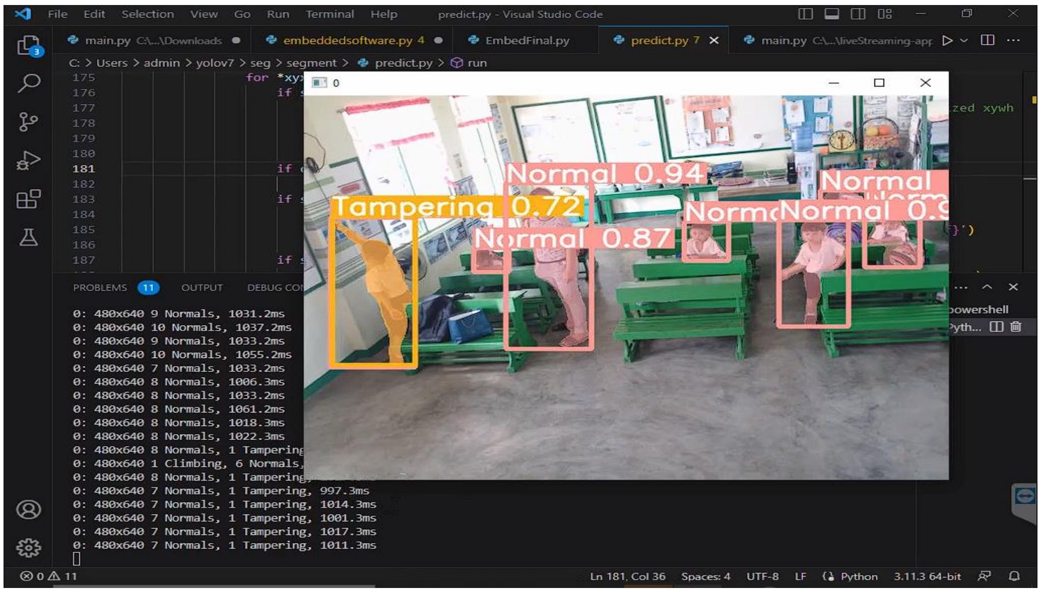


Figure 19. Detection of Tampering Outlet

Figure 19 shows the sample detection of tampering class in a frame upon testing in the classroom environment of Manalupang San Vicente Elementary School. This displays the live stream of kids which undergo instance segmentation and classification of running with a confidence score of 72% and normal classification that plays around 87% to 90% or above.

CHAPTER V

RESULTS AND DISCUSSION

This chapter presents the discussion of results, and summary of findings upon the completion of the project.

5.1 DISCUSSION OF RESULTS

5.1.1 Dataset Collection

This successfully gathered an image dataset of various safety hazards from the kindergarten environment in collaboration with the Principal and Kindergarten Teacher of Manalupang - San Vicente Elementary School. Through an agreement with the school, these were granted access to the kindergarten classroom and the students, ensuring their direct involvement and benefit from the project.

Throughout the project, this recorded class session and extracted images that captured hazardous situations, including instances of climbing on chairs and tables, running, and tampering with electrical outlets. These images were carefully organized and labeled to create a comprehensive dataset that serves as a valuable resource for developing safety awareness programs in kindergartens. By involving the Principal and Kindergarten Teacher in the project, it ensured that the dataset's creation aligned with the specific needs and concerns of the school community, fostering a collaborative and impactful approach to safety education.

Table 10. Sample Datasets from Kindergarten Classroom

Running & Normal	
Climbing	
Tampering	

Table 10 shows the sample images from the recorded class session. The gathered dataset serves as valuable input for the development and training of intelligent detection models aimed at enhancing safety measures within the kindergarten environment.

5.1.2 Fine Tuning and Comparison of Fine Tuned Pretrained Models

Upon fine tuning of different models, specifically Mask R-CNN, YOLOv7 2D Pose and instance segmentation, the performance evaluation was conducted to determine the most suitable model for detection of safety hazard.

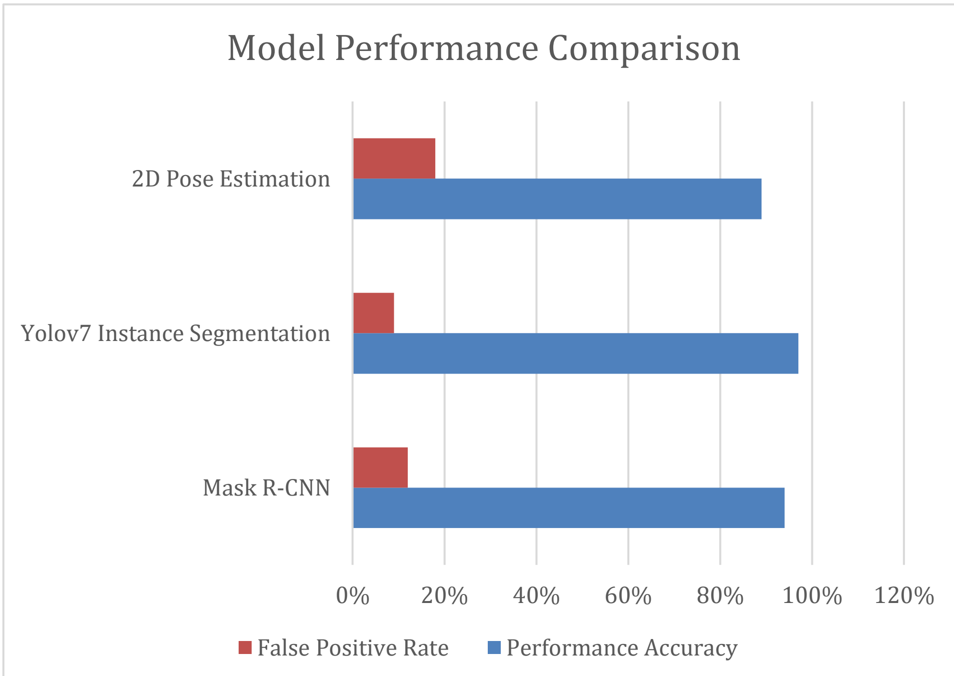


Figure 20. Performance Evaluation of The Three Model

Figure 20 illustrates the Mask R-CNN model which achieved an accuracy of 89% upon testing from the unique dataset separated from training. Also, this model attains a 12% false positive rate indicating a few false predictions from the testing and validation process. For Yolov7 Instance Segmentation, the table shows that it accomplishes a 97% accuracy compared to the other and garner a 9% false positive rate which was the lowest of the three models in comparison.

Lastly, for the 2D Pose Estimation, shown above was an accuracy of 89% which was the lowest among the models tested. Unlike the other models, the false positive rate of this model was significantly higher than the two mentioned above indicating a large number of false predictions.

5.1.3 Integration of Developed Software and Hardware System

After collecting the dataset and finding the most suitable model, which was the Yolov7 Instance Segmentation, this move on to integration of the software and hardware.

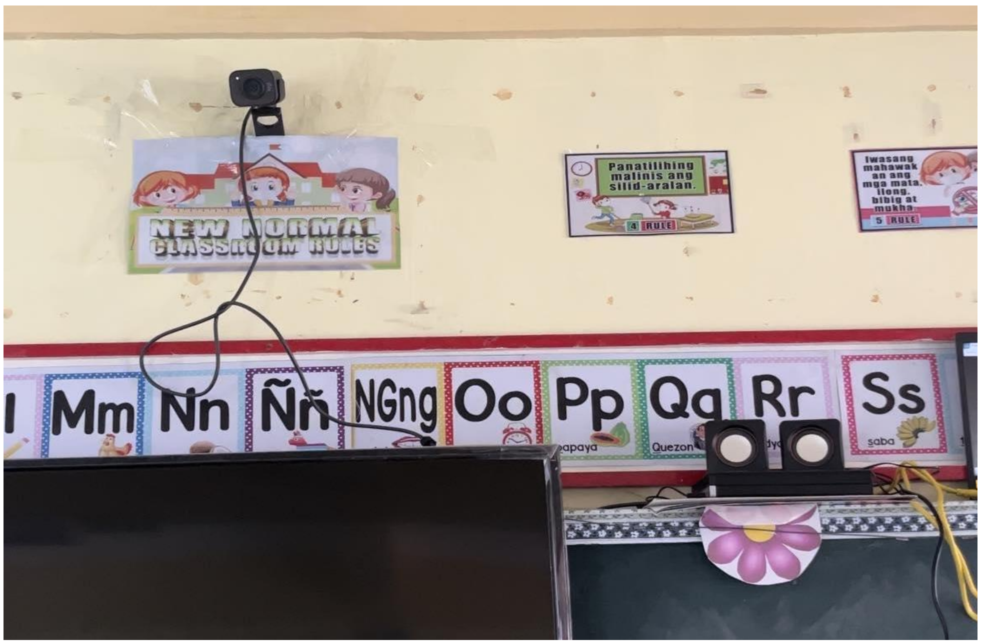
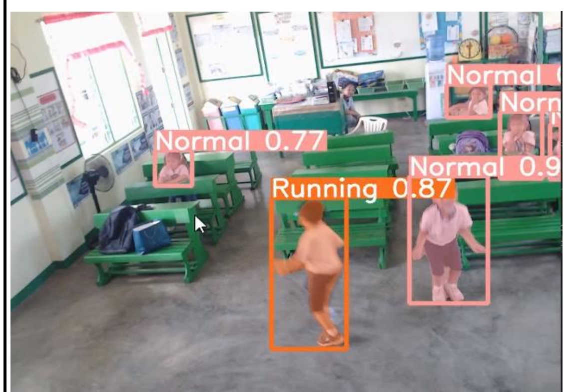


Figure 21. Output Device with the Integrated Software

The output device was presented in figure 21 which includes a camera as input, which was processed by an integrated software system consisting of a mini-PC and software code. The segmented output was analyzed, and an appropriate alarm was selected based on the detected class, triggering a speaker. A live stream of the students can be accessed through TeamViewer, allowing the teacher to view the classroom remotely and receive alarms.

Table 11. Sample Output from the Recorded Testing

Climbing	
Running	
Tampering	

Shown in the table 11 was the sample output for each of the class upon conducting live testing in Manalupang San Vicente Elementary School with the output device. It can be glanced from the sample, climbing, running, tampering and normal were detected with different confidence score.

$$Av = \frac{\textit{Sum of confidence score per class}}{\textit{Total num of prediction per class}} \quad \text{eq (2)}$$

Throughout the live testing, different confidence scores were recorded and to evaluate the output device, it takes the average confidence score for each class predictions and this was achieved through equation 2.

Table 12. Average Confidence Score Per Classification

Class	Average Confidence Scores
Climbing	94%
Running	90%
Tampering	86%

Table 12 presents the average confidence scores per classification upon live testing in the Kindergarten environment. Based on the table, the system has a higher confidence score in terms of the Climbing class followed by the running and tampering. The system has 94% confidence in terms of classifying Climbing while 90% for Running and 86% for Tampering.

5.2 SUMMARY OF FINDINGS

The following are the findings based on the results obtained from the design and development of an IoT-based alarm and monitoring system for kindergarten classroom of Manalupang-San Vicente Elementary School, that detects hazardous situations present in the classroom.

This successfully gathered an image dataset of various safety hazards in a kindergarten environment. Specifically climbing, running, and tampering outlet. The dataset was collected in collaboration with the Principal and Kindergarten Teacher of Manalupang - San Vicente Elementary School.

After collecting the dataset, the models were trained using Google Colab and achieved 97% accuracy and 9% false positive rate in terms of yolov7 instance segmentation followed by Mask R-CNN with an accuracy of 94% and 12% false positive. Lastly, the 2D pose with an accuracy of 89% and 18% false positive rate. For this, yolov7 instance segmentation has the highest accuracy and lower false positive rate which leads the researchers in using this model for detection of safety hazards.

The system integrated the trained model with a camera, mini-PC, and software code. The segmented output was analyzed, and an appropriate alarm was triggered based on the detected class, alerting the teacher. The teacher could also view a live stream of the classroom remotely through TeamViewer.

The developed software and hardware system, along with the trained YOLOv7 model, successfully detected safety hazards in the kindergarten environment. The system demonstrated a high confidence score in detecting instances of running, climbing, and tampering with outlets, which were prevalent hazardous behaviors exhibited by the students. The integration of the system with live testing proved to be highly successful, providing a reliable means of enhancing safety measures within the kindergarten classroom.

CONCLUSIONS

This successfully collected an image dataset of various safety hazards present in a typical kindergarten environment. This dataset was obtained through collaboration with the Principal and Kindergarten Teacher of Manalupang-San Vicente Elementary School. The hazards included are climbing, running, and tampering with outlets, which are common risky behaviors exhibited by young children.

Secondly, among the models tested, the yolov7 instance segmentation model achieved the highest accuracy rate of 97% with a low false positive rate of 9%. This model outperformed the Mask R-CNN and 2D pose estimation models, which achieved accuracies of 94% and 89% respectively. Therefore, the yolov7 model was selected as the most suitable for detecting safety hazards in the classroom.

Thirdly, the trained model was successfully integrated into a comprehensive system comprising a camera, mini-PC, and software code. This integrated system analyzed the segmented output from the model and triggered appropriate alarms based on the detected hazard class, immediately alerting the teacher. Furthermore, the system allowed the teacher to remotely access a live stream of the classroom using TeamViewer, enhancing their ability to monitor and respond to potential risks.

Finally, the developed software and hardware system, in combination with the yolov7 model, effectively detected safety hazards in the kindergarten environment. The system demonstrated a high level of accuracy in identifying instances of running, climbing, and tampering with outlets, which are prevalent and potentially dangerous behaviors exhibited by young students. The successful integration of the system with live testing proved its reliability and provided an efficient means of enhancing safety measures within the kindergarten classroom.

RECOMMENDATION FOR FUTURE WORK

The developed device successfully detects trained safety hazards with an impressive accuracy of 97%. It has proven to be an effective assistant for teachers, although the researchers believe that improvement would still manifest. Several recommendations can enhance the device's functionality:

1. Include a wider range of safety hazards in the training dataset to enhance detection capabilities.
2. Utilize GPU technology to accelerate the hazard classification process and improve the device's speed and efficiency.
3. Explore the application of the device in different settings, such as homes or controlled environments, to extend its usefulness in detecting safety hazards.
4. Incorporate a wide-angle camera into the device to maximize its coverage and ensure comprehensive monitoring of the surroundings.
5. Investigate alternative algorithms in addition to YoloV7 to assess their suitability and potentially enhance the device's performance.

BIBLIOGRAPHY

Adorama Learning Center (2021). What is the Best Camera for Live Streaming Video? - 42 West. [online] 13 Apr. Available at: https://www.adorama.com/alc/what-is-the-best-camera-for-live-streaming-video/.

Anushkannan, N.K., Kumbhar, V.R., Maddila, S.K., Kolli, C.S., Vidhya, B. and Vidhya, R.G. (2022). *YOLO Algorithm for Helmet Detection in Industries for Safety Purpose*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICOSEC54921.2022.9952154.

Arief, L., Tantowi, A., Novani, N.P. and Sundara, T.A. (2020). Implementation of YOLO and Smoke Sensor for Automating Public Service Announcement of Cigarette's Hazard in Public Facilities. *2020 International Conference on Information Technology Systems and Innovation (ICITSI)*. [online] doi:https://doi.org/10.1109/ICITSI50517.2020.9264972.

Arora, A., Vijayvargiya, A., Kumar, R. and Tiwari, M. (2021). *Machine Learning based Risk Classification of Musculoskeletal Disorder among the Garment Industry Operators*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICIRCA51532.2021.9544820.

Ashtekar, S., Dhalwar, S. and Pasupathy, A. (2021). *Computer Vision Based Vulnerable Road Users Hand Signal Recognition*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICIRCA51532.2021.9545005.

Bolya, D., Fanyi, C., Yong, X. and Lee, J. (n.d.). YOLACT Real-time Instance Segmentation. [online] Available at: https://openaccess.thecvf.com/content_ICCV_2019/papers/Bolya_YOLACT_Real-Time_Instance_Segmentation_ICCV_2019_paper.pdf

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E. and Sheikh, Y. (2019). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. arXiv:1812.08008 [cs]. [online] Available at: https://arxiv.org/abs/1812.08008.

Chang, R., Zhou, S., Zhang, Y., Zhang, N., Zhou, C. and Li, M. (2023). Research on Insulator Defect Detection Based on Improved YOLOv7 and Multi-UAV Cooperative System. *Coatings*, [online] 13(5), p.880. doi:https://doi.org/10.3390/coatings13050880.

Chen, C. and Da, S. (2022). *Research on Ship's Officer Behavior Identification Based on Mask R-CNN*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICRCV55858.2022.9953184.

Chen, P., Wu, Y., Qin, Y. and Yang, H. (2022). All-in-One YOLO Architecture for safety Hazard Detection of Environment along High-Speed Railway. doi:https://doi.org/10.1109/phm-yantai55411.2022.9941973.

Fang, W., Zhong, B., Zhao, N., Love, P.E.D., 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*, [online] 39, pp.170–177. doi:https://doi.org/10.1016/j.aei.2018.12.005.

Gan, H., Ou, M., Li, C., Wang, X., Guo, J., Mao, A., Camila Ceballos, M., Parsons, T.D., Liu, K. and Xue, Y. (2022). Automated detection and analysis of piglet suckling behaviour using high-accuracy amodal instance segmentation. *Computers and Electronics in Agriculture*, [online] 199, p.107162. doi:https://doi.org/10.1016/j.compag.2022.107162.

Hermans, A., Chen, L.-C., Papandreou, G., Schroff, F., Wang, P. and Adam, H. (2018). MaskLab: Instance Segmentation by Refining Object Detection with Semantic and Direction Features. [online] arXiv.org. doi:https://doi.org/10.48550/arXiv.1712.04837.

Jeelani, I., Asadi, K., Ramshankar, H., Han, K. and Albert, A. (2021). Real-time vision-based worker localization & hazard detection for construction. *Automation in Construction*, 121, p.103448. doi:https://doi.org/10.1016/j.autcon.2020.103448.

Jiang, K., Xie, T., Yan, R., Wen, X., Li, D., Jiang, H., Jiang, N., Feng, L., Duan, X. and Wang, J. (2022). An Attention Mechanism-Improved YOLOv7 Object Detection Algorithm for Hemp Duck Count Estimation. *Agriculture*, 12(10), p.1659. doi:https://doi.org/10.3390/agriculture12101659.

Kaya, Ö., Çodur, M.Y. and Mustafaraj, E. (2023). Automatic Detection of Pedestrian Crosswalk with Faster R-CNN and YOLOv7. *Buildings*, 13(4), p.1070. doi:https://doi.org/10.3390/buildings13041070.

Khan, N., Saleem, M.R., Lee, D., Park, M.-W. and Park, C. (2021). Utilizing safety rule correlation for mobile scaffolds monitoring leveraging deep convolution neural networks. *Computers in Industry*, 129, p.103448. doi:https://doi.org/10.1016/j.compind.2021.103448.

Kim, T.T., Zohdy, M.A. and Barker, M.P. (2020). Applying Pose Estimation to Predict Amateur Golf Swing Performance Using Edge Processing. IEEE Access, 8, pp.143769–143776. doi:https://doi.org/10.1109/access.2020.3014186.

Kurtanović J. (2021). Deep Learning – Instance Segmentation | Serengeti. [online] Available at: https://serengetitech.com/tech/deep-learning-instance-segmentation/.

Le, T.D., Huynh, D. and Pham, H.P. (2018). Efficient Human-Robot Interaction using

Deep Learning with Mask R-CNN: Detection, Recognition, Tracking and Segmentation. doi:https://doi.org/10.1109/icarcv.2018.8581081.

Luvizon, D., Picard, D. and Tabia, H. (n.d.). 2D/3D Pose Estimation and Action Recognition using Multitask Deep Learning. [online] Available at: https://openaccess.thecvf.com/content_cvpr_2018/papers/Luvizon_2D3D_Pose_Estimation_CVPR_2018_paper.pdf.

McNally, W., Vats, K., Wong, A. and McPhee, J. (2021). EvoPose2D: Pushing the Boundaries of 2D Human Pose Estimation using Accelerated Neuroevolution with Weight Transfer. IEEE Access, [online] 9, pp.139403–139414. doi:https://doi.org/10.1109/ACCESS.2021.3118207.

Mobsite, S., Alaoui, N. and Boulmalf, M. (2020). *A framework for elders fall detection using deep learning*. [online] IEEE Xplore. doi:https://doi.org/10.1109/CiSt49399.2021.9357184.

Noe, S.M., Zin, T.T., Tin, P. and Kobayashi, I. (2022). Automatic detection and tracking of mounting behavior in cattle using a deep learning-based instance segmentation model. *Int. J. Innov. Comput. Inf. Control*, 18, pp.211–220.

Osokin, D. (2018). Real-time 2D Multi-Person Pose Estimation on CPU: Lightweight OpenPose. [online] Available at: https://arxiv.org/pdf/1811.12004.pdf.

Peng, S., Jiang, W., Pi, H., Li, X., Bao, H. and Zhou, X. (2020). Deep Snake for Real-Time Instance Segmentation. [online] arXiv.org. doi:https://doi.org/10.48550/arXiv.2001.01629.

Poudel, R., Lima, L. and Andrade, F. (2023). *A Novel Framework To Evaluate and Train Object Detection Models for Real-Time Victims Search and Rescue at Sea With Autonomous Unmanned Aerial Systems Using High-Fidelity Dynamic Marine Simulation Environment*. [online] openaccess.thecvf.com. Available at: https://openaccess.thecvf.com/content/WACV2023W/MaCVi/html/Poudel_A_Novel_Framework_To_Evaluate_and_Train_Object_Detection_Models_WACVW_2023_paper.html [Accessed 6 Jun. 2023].

Raaj, Y., Idrees, H., Hidalgo, G. and Sheikh, Y. (2019). Efficient Online Multi-Person 2D Pose Tracking with Recurrent Spatio-Temporal Affinity Fields. [online] arXiv.org. doi:https://doi.org/10.48550/arXiv.1811.11975.

Reddy, E.S.T.K. and V, R. (2022). Pothole Detection using CNN and YOLO v7 Algorithm. *2022 6th International Conference on Electronics, Communication and Aerospace Technology*. doi:https://doi.org/10.1109/iceca55336.2022.10009324.

Sun, G. and Wang, Z. (2020). Fall detection algorithm for the elderly based on human posture estimation. doi:https://doi.org/10.1109/ipec49694.2020.9114962.

Tompson, J., Papandreou, G., Zhu, T., Chen, L.-C., Gidaris, S., and Murphy, K. (2018).

PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model. [online] arXiv.org. doi:https://doi.org/10.48550/arXiv.1803.08225.

Torres, M. (2023). What Is Fine-Tuning and How Does It Work in Neural Networks? [online] blog.pangeanic.com. Available at: https://blog.pangeanic.com/what-is-fine-tuning [Accessed 16 Jun. 2023].

Wang, Y., Li, M., Cai, H., Chen, W.-M. and Han, S. (2022). Lite Pose: Efficient Architecture Design for 2D Human Pose Estimation. [online] openaccess.thecvf.com. Available at: https://openaccess.thecvf.com/content/CVPR2022/html/Wang_Lite_Pose_Efficient_Architecture_Design_for_2D_Human_Pose_Estimation_CVPR_2022_paper.html.

Wei, F., Sun, X., Li, H., Wang, J. and Lin, S. (2020). Point-Set Anchors for Object Detection, Instance Segmentation and Pose Estimation. [online] arXiv.org. doi:https://doi.org/10.48550/arXiv.2007.02846.

What is the Best Camera for Live Streaming Video? - 42 West. (2021). 42 West, the Adorama Learning Center. [online] 13 Apr. Available at: https://www.adorama.com/alc/what-is-the-best-camera-for-live-streaming-video/.

Wu, Y., Meng, F., Qin, Y., Qian, Y., Xu, F. and Jia, L. (2023). UAV imagery based potential safety hazard evaluation for high-speed railroad using Real-time instance segmentation. *Advanced Engineering Informatics*, 55, p.101819. doi:https://doi.org/10.1016/j.aei.2022.101819.

Xiao, Y., Chang, A., Wang, Y., Huang, Y., Yu, J. and Huo, L. (2022a). *Real-time Object Detection for Substation Security Early-warning with Deep Neural Network based on YOLO-V5*. [online] IEEE Xplore. doi:https://doi.org/10.1109/GlobConET53749.2022.9872338.

Xiao, Y., Wang, Y., Li, W., Sun, M., Shen, X. and Luo, Z. (2022b). Monitoring the Abnormal Human Behaviors in Substations based on Probabilistic Behaviours Prediction and YOLO-V5. doi:https://doi.org/10.1109/acpee53904.2022.9783954.

Xie, E., Sun, P., Song, X., Wang, W., Liu, X., Liang, D., Shen, C. and Luo, P. (n.d.). PolarMask: Single Shot Instance Segmentation with Polar Representation. [online] Available at: https://openaccess.thecvf.com/content_CVPR_2020/papers/Xie_PolarMask_Single_Shot_Instance_Segmentation_With_Polar_Representation_CVPR_2020_paper.pdf

Xiong, R., Song, Y., Li, H. and Wang, Y. (2019). Onsite video mining for construction hazards identification with visual relationships. *Advanced Engineering Informatics*, [online] 42, p.100966. doi:https://doi.org/10.1016/j.aei.2019.100966.

Xu, J., Wu, Q., Zhang, J. and Tait, A. (2021). *AUTOMATIC SHEEP BEHAVIOUR ANALYSIS USING MASK R-CNN*. [online] IEEE Xplore. doi:https://doi.org/10.1109/DICTA52665.2021.9647101.

Yang, Z., Yuan, Y., Zhang, M., Zhao, X., Zhang, Y. and Tian, B. (2019). Safety Distance Identification for Crane Drivers Based on Mask R-CNN. *Sensors*, 19(12), p.2789. doi:https://doi.org/10.3390/s19122789.

Zhang, S.-H., Li, R., Dong, X., Rosin, P., Cai, Z., Han, X., Yang, D., Huang, H. and Hu, S.-M. (n.d.). Pose2Seg: Detection Free Human Instance Segmentation. [online] Available at: https://openaccess.thecvf.com/content_CVPR_2019/papers/Zhang_Pose2Seg_Detection_Free_Human_Instance_Segmentation_CVPR_2019_paper.pdf

Zhang, Y. and Zhou, Y. (2021). *YOLOv5 Based Pedestrian Safety Detection in Underground Coal Mines*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ROBIO54168.2021.9739594.

Zhao, H., Zhang, H. and Zhao, Y. (n.d.). YOLOv7-sea: Object Detection of Maritime UAV Images based on Improved YOLOv7. [online] Available at: https://openaccess.thecvf.com/content/WACV2023W/MaCVi/papers/Zhao_YOLOv7-Sea_Object_Detection_of_Maritime_UAV_Images_Based_on_Improved_WACVW_2023_paper.pdf.

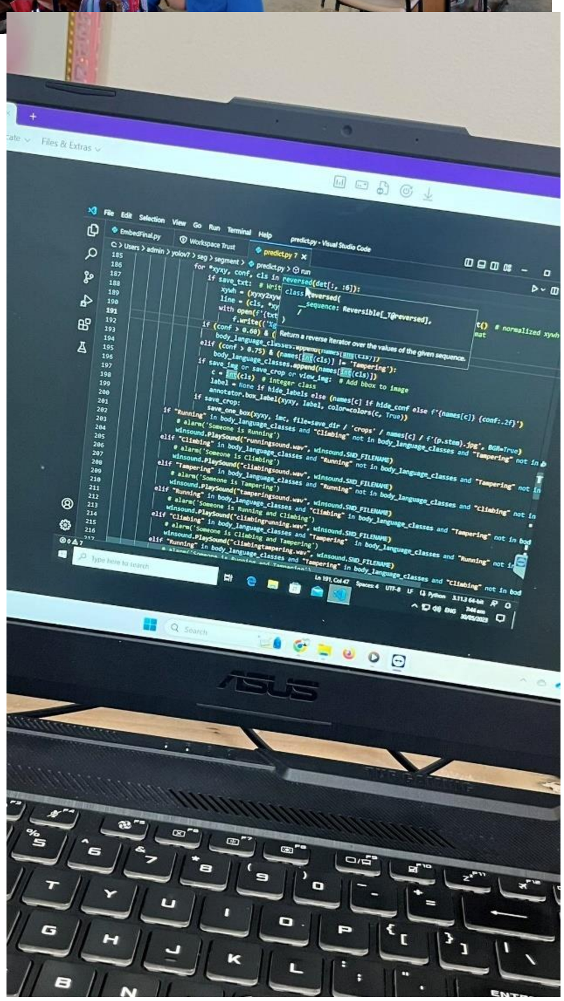
Zherdeva, L., Zherdev, D. and Nikonorov, A. (2021). *Prediction of human behavior with synthetic data*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ITNT52450.2021.9649073.

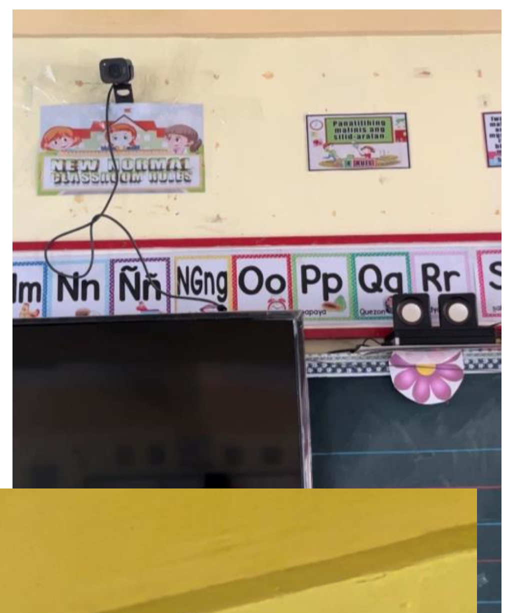
Zhu, P., Chen, B., Liu, B., Qi, Z., Wang, S. and Wang, L. (2023). Object Detection for Hazardous Material Vehicles Based on Improved YOLOv5 Algorithm. *Electronics*, [online] 12(5), p.1257. doi:https://doi.org/10.3390/electronics12051257.

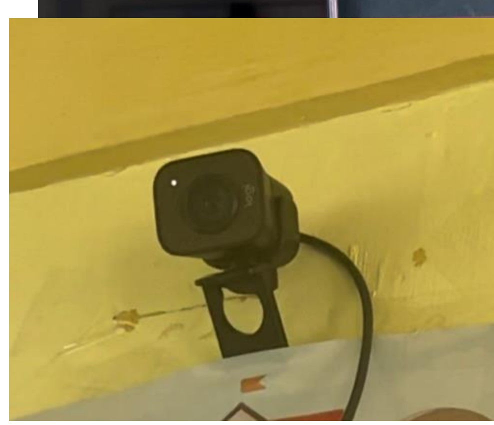
Zi, X., Chaturvedi, K., Braytee, A., Li, J. and Prasad, M. (2023). Detecting Human Falls in Poor Lighting: Object Detection and Tracking Approach for Indoor Safety. *Electronics*, [online] 12(5), p.1259. doi:https://doi.org/10.3390/electronics12051259.

Appendix A

Documentation







Appendix B

Code

```
import os
HOME = '(os.getcwd)/yolov7'
import subprocess
command = "python yolov7/seg/segment/predict.py --weights yolov7/seg/runs/train-
seg/custom5/weights/best.pt --source 0"
try:
    subprocess.run(command, shell=True, check=True)
except subprocess.CalledProcessError as e:
    print(f"Command failed with return code {e.returncode}")
import argparse
import os
import platform
import sys
from pathlib import Path

import torch
import torch.backends.cudnn as cudnn
import winsound
from tracemalloc import stop
import requests
```

```
import socket


FILE = Path(__file__).resolve()
ROOT = FILE.parents[1]  # YOLOv5 root directory
if str(ROOT) not in sys.path:
    sys.path.append(str(ROOT))  # add ROOT to PATH
ROOT = Path(os.path.relpath(ROOT, Path.cwd()))  # relative


from models.common import DetectMultiBackend
from utils.dataloaders import IMG_FORMATS, VID_FORMATS, LoadImages,
LoadStreams
from utils.general import (LOGGER, Profile, check_file, check_img_size,
check_imshow, check_requirements, colorstr, cv2,
                           increment_path, non_max_suppression, print_args, scale_coords,
strip_optimizer, xyxy2xywh)
from utils.plots import Annotator, colors, save_one_box
from utils.segment.general import process_mask, scale_masks
from utils.segment.plots import plot_masks
from utils.torch_utils import select_device, smart_inference_mode



@smart_inference_mode()
def run(
```

```
weights=ROOT / 'yolov5s-seg.pt',  # model.pt path(s)
source=ROOT / 'data/images',  # file/dir/URL/glob, 0 for webcam
data=ROOT / 'data/coco128.yaml',  # dataset.yaml path
imgsz=(480, 848),  # inference size (height, width)
conf_thres=0.50,  # confidence threshold
iou_thres=0.45,  # NMS IOU threshold
max_det=1000,  # maximum detections per image
device='',  # cuda device, i.e. 0 or 0,1,2,3 or cpu
view_img=True,  # show results
save_txt=False,  # save results to *.txt
save_conf=False,  # save confidences in --save-txt labels
save_crop=False,  # save cropped prediction boxes
nosave=False,  # do not save images/videos
classes=None,  # filter by class: --class 0, or --class 0 2 3
agnostic_nms=False,  # class-agnostic NMS
augment=True,  # augmented inference
visualize=True,  # visualize features
update=False,  # update all models
project=ROOT / 'runs/predict-seg',  # save results to project/name
name='exp',  # save results to project/name
exist_ok=False,  # existing project/name ok, do not increment
line_thickness=3,  # bounding box thickness (pixels)
hide_labels=False,  # hide labels
```

```
        hide_conf=False,  # hide confidences
        half=False,  # use FP16 half-precision inference
        dnn=False,  # use OpenCV DNN for ONNX inference
):
    source = str(source)
    save_img = not nosave and not source.endswith('.txt')  # save inference images
    is_file = Path(source).suffix[1:] in (IMG_FORMATS + VID_FORMATS)
    is_url = source.lower().startswith(('rtsp://', 'rtmp://', 'http://', 'https://'))
    webcam = source.isnumeric() or source.endswith('.txt') or (is_url and not is_file)
    if is_url and is_file:
        source = check_file(source)  # download

    # Directories
    save_dir = increment_path(Path(project) / name, exist_ok=exist_ok)  # increment run
    (save_dir / 'labels' if save_txt else save_dir).mkdir(parents=True, exist_ok=True)  #
make dir

    # Load model
    device = select_device(device)
    model = DetectMultiBackend(weights, device=device, dnn=dnn, data=data, fp16=half)
    stride, names, pt = model.stride, model.names, model.pt
    imgsz = check_img_size(imgsz, s=stride)  # check image size
    # imgsz = (480, 640)
```

```
# Dataloader
if webcam:
    view_img = check_imshow()
    cudnn.benchmark = True  # set True to speed up constant image size inference
    dataset = LoadStreams(source, img_size=imgsz, stride=stride, auto=pt)
    bs = len(dataset)  # batch_size
else:
    dataset = LoadImages(source, img_size=imgsz, stride=stride, auto=pt)
    bs = 1  # batch_size
vid_path, vid_writer = [None] * bs, [None] * bs
# Run inference
model.warmup(imgsz=(1 if pt else bs, 3, *imgsz))  # warmup
seen, windows, dt = 0, [], (Profile(), Profile(), Profile())
for path, im, im0s, vid_cap, s in dataset:
    import torch
    from torchvision import transforms

    with dt[0]:
        # im0s = cv2.resize(im0s, (640, 480), interpolation=cv2.INTER_LINEAR)
        im = torch.from_numpy(im).to(device)
        #         # Create a resize transform
        # resize_transform = transforms.Resize((480, 640), antialias=True)
```

```
        # # Apply the resize transform to the input image tensor
        # resized_im_tensor = resize_transform(im)

        # # Move the tensor to the desired device (e.g., GPU)
        # im = resized_im_tensor.to(device)

        im = im.half() if model.fp16 else im.float()  # uint8 to fp16/32
        im /= 255  # 0 - 255 to 0.0 - 1.0
        if len(im.shape) == 3:
            im = im[None]  # expand for batc
    # Inference
    with dt[1]:
        visualize = increment_path(save_dir / Path(path).stem, mkdir=True) if visualize
else False
        pred, out = model(im, augment=augment, visualize=visualize)
        proto = out[1]
    # print(im)
    # NMS
    with dt[2]:
        pred = non_max_suppression(pred, conf_thres, iou_thres, classes, agnostic_nms,
max_det=max_det, nm=32)
```

```
        # Second-stage classifier (optional)
        # pred = utils.general.apply_classifier(pred, classifier_model, im, im0s)

        # Process predictions
        for i, det in enumerate(pred):  # per image
            seen += 1
            if webcam:  # batch_size >= 1
                p, im0, frame = path[i], im0s[i].copy(), dataset.count
                s += f'{i}: '
            else:
                p, im0, frame = path, im0s.copy(), getattr(dataset, 'frame', 0)

            p = Path(p)  # to Path
            save_path = str(save_dir / p.name)  # im.jpg
            txt_path = str(save_dir / 'labels' / p.stem) + ('' if dataset.mode == 'image' else
f'_{frame}')  # im.txt
            s += '%gx%g ' % im.shape[2:]  # print string
            gn = torch.tensor(im0.shape)[[1, 0, 1, 0]]  # normalization gain whwh
            imc = im0.copy() if save_crop else im0  # for save_crop
            annotator = Annotator(im0, line_width=line_thickness, example=str(names))
            body_language_classes = []
            if len(det):
                masks = process_mask(proto[i], det[:, 6:], det[:, :4], im.shape[2:],
```

upsample=True) # HWC

Rescale boxes from img_size to im0 size

det[:, :4] = scale_coords(im.shape[2:], det[:, :4], im0.shape).round()

Print results

for c in det[:, 5].unique():

n = (det[:, 5] == c).sum() # detections per class

s += f"{n} {names[int(c)]}{'s' * (n > 1)}, " # add to string

Mask plotting ---
----

mcolors = [colors(int(cls), True) for cls in det[:, 5]]

# im_masks = plot_masks(im[i], masks, mcolors) # image with masks
shape(imh,imw,3)

annotator.im = scale_masks(im.shape[2:], im_masks, im0.shape) # scale to
original h, w

Mask plotting ---

Write results

for *xyxy, conf, cls in reversed(det[:, :6]):

if save_txt: # Write to file

```
                    xywh = (xyxy2xywh(torch.tensor(xyxy).view(1, 4)) / gn).view(-1).tolist()
# normalized xywh
                    line = (cls, *xywh, conf) if save_conf else (cls, *xywh)  # label format
                    with open(f'{txt_path}.txt', 'a') as f:
                        f.write(('%g ' * len(line)).rstrip() % line + '\n')
                if (conf > 0.60) & (names[int(cls)] == 'Tampering'):
                    body_language_classes.append(names[int(cls)])
                elif (conf > 0.75) & (names[int(cls)] != 'Tampering'):
                    body_language_classes.append(names[int(cls)])
                if save_img or save_crop or view_img:  # Add bbox to image
                    c = int(cls)  # integer class
                    label = None if hide_labels else (names[c] if hide_conf else f'{names[c]}
{conf:.2f}')
                    annotator.box_label(xyxy, label, color=colors(c, True))
                if save_crop:
                    save_one_box(xyxy, imc, file=save_dir / 'crops' / names[c] /
f'{p.stem}.jpg', BGR=True)
        if "Running" in body_language_classes and "Climbing" not in
body_language_classes and "Tampering" not in body_language_classes:
            # alarm('Someone is Running')
            winsound.PlaySound("runningsound.wav", winsound.SND_FILENAME)
        elif "Climbing" in body_language_classes and "Running" not in
body_language_classes and "Tampering" not in body_language_classes:
```

```
        # alarm('Someone is Climbing')
        winsound.PlaySound("climbingsound.wav", winsound.SND_FILENAME)
    elif "Tampering" in body_language_classes and "Running" not in
body_language_classes and "Climbing" not in body_language_classes:
        # alarm('Someone is Tampering')
        winsound.PlaySound("tamperingsound.wav", winsound.SND_FILENAME)
    elif "Running" in body_language_classes and "Climbing" in
body_language_classes and "Tampering" not in body_language_classes:
        # alarm('Someone is Running and Climbing')
        winsound.PlaySound("climbingrunning.wav", winsound.SND_FILENAME)
    elif "Climbing" in body_language_classes and "Tampering"  in
body_language_classes and "Running" not in body_language_classes:
        # alarm('Someone is Climbing and Tampering')
        winsound.PlaySound("climbingtampering.wav", winsound.SND_FILENAME)
    elif "Running" in body_language_classes and "Tampering" in
body_language_classes and "Climbing" not in body_language_classes:
        # alarm('Someone is Running and Tampering')
        winsound.PlaySound("runningtampering.wav", winsound.SND_FILENAME)
    elif "Running" in body_language_classes and "Tampering" in
body_language_classes and "Climbing" in body_language_classes:
        # alarm('Someone is Running, Climbing and Tampering')
        winsound.PlaySound("all.wav", winsound.SND_FILENAME)
    # If the body language class is not one that triggers the alarm, turn it off
```

```
            elif "Running" in body_language_classes and "Tampering" not in
body_language_classes and "Running" not in body_language_classes:
                stop()
            # Stream results
            # im0 = annotator.result()
            if view_img:
                if platform.system() == 'Linux' and p not in windows:
                    windows.append(p)
                    cv2.namedWindow(str(p), cv2.WINDOW_NORMAL |
cv2.WINDOW_KEEPRATIO)  # allow window resize (Linux)
                    # cv2.resizeWindow(str(p), im0.shape[1], im0.shape[0])
                    cv2.resizeWindow(str(p), 1920, 1080)
                cv2.imshow(str(p), im0)
                cv2.waitKey(1)  # 1 millisecond

            # Save results (image with detections)
            if save_img:
                if dataset.mode == 'image':
                    cv2.imwrite(save_path, im0)
                else:  # 'video' or 'stream'
                    if vid_path[i] != save_path:  # new video
                        vid_path[i] = save_path
                        if isinstance(vid_writer[i], cv2.VideoWriter):
```

```
                    vid_writer[i].release()  # release previous video writer
                if vid_cap:  # video
                    fps = vid_cap.get(cv2.CAP_PROP_FPS)
                    # w = 1920
                    # h = 1080
                    w = int(vid_cap.get(cv2.CAP_PROP_FRAME_WIDTH))
                    h = int(vid_cap.get(cv2.CAP_PROP_FRAME_HEIGHT))
                else:  # stream
                    fps, w, h = 30, im0.shape[1], im0.shape[0]
                save_path = str(Path(save_path).with_suffix('.mp4'))  # force *.mp4 suffix
on results videos
                vid_writer[i] = cv2.VideoWriter(save_path,
cv2.VideoWriter_fourcc(*'mp4v'), fps, (w, h))
            vid_writer[i].write(im0)


        # Print time (inference-only)
        LOGGER.info(f"{s}{'' if len(det) else '(no detections), '}{dt[1].dt * 1E3:.1f}ms")


    # Print results
    t = tuple(x.t / seen * 1E3 for x in dt)  # speeds per image
    LOGGER.info(f'Speed: %.1fms pre-process, %.1fms inference, %.1fms NMS per
image at shape {(1, 3, *imgsz)}' % t)
    if save_txt or save_img:
```

```
        s = f"\n{len(list(save_dir.glob('labels/*.txt')))} labels saved to {save_dir / 'labels'}"
if save_txt else ''
        LOGGER.info(f"Results saved to {colorstr('bold', save_dir)}{s}")
    if update:
        strip_optimizer(weights[0])  # update model (to fix SourceChangeWarning)


def parse_opt():
    parser = argparse.ArgumentParser()
    parser.add_argument('--weights', nargs='+', type=str, default=ROOT / 'yolov5s-seg.pt',
help='model path(s)')
    parser.add_argument('--source', type=str, default=ROOT / 'data/images',
help='file/dir/URL/glob, 0 for webcam')
    parser.add_argument('--data', type=str, default=ROOT / 'data/coco128.yaml',
help='(optional) dataset.yaml path')
    parser.add_argument('--imgsz', '--img', '--img-size', nargs='+', type=int, default=[250],
help='inference size h,w')
    parser.add_argument('--conf-thres', type=float, default=0.35, help='confidence
threshold')
    parser.add_argument('--iou-thres', type=float, default=0.45, help='NMS IoU threshold')
    parser.add_argument('--max-det', type=int, default=1000, help='maximum detections
per image')
    parser.add_argument('--device', default='', help='cuda device, i.e. 0 or 0,1,2,3 or cpu')
```

parser.add_argument('--view-img', action='store_false', help='show results')

parser.add_argument('--save-txt', action='store_false', help='save results to *.txt')

parser.add_argument('--save-conf', action='store_false', help='save confidences in --save-txt labels')

parser.add_argument('--save-crop', action='store_false', help='save cropped prediction boxes')

parser.add_argument('--nosave', action='store_true', help='do not save images/videos')

parser.add_argument('--classes', nargs='+', type=int, help='filter by class: --classes 0, or --classes 0 2 3')

parser.add_argument('--agnostic-nms', action='store_true', help='class-agnostic NMS')

parser.add_argument('--augment', action='store_true', help='augmented inference')

parser.add_argument('--visualize', action='store_true', help='visualize features')

parser.add_argument('--update', action='store_false', help='update all models')

parser.add_argument('--project', default=ROOT / 'runs/predict-seg', help='save results to project/name')

parser.add_argument('--name', default='exp', help='save results to project/name')

parser.add_argument('--exist-ok', action='store_true', help='existing project/name ok, do not increment')

parser.add_argument('--line-thickness', default=3, type=int, help='bounding box thickness (pixels)')

parser.add_argument('--hide-labels', default=True, action='store_true', help='hide labels')

parser.add_argument('--hide-conf', default=True, action='store_true', help='hide

```
confidences')
    parser.add_argument('--half', action='store_true', help='use FP16 half-precision
inference')
    parser.add_argument('--dnn', action='store_true', help='use OpenCV DNN for ONNX
inference')
    opt = parser.parse_args()
    opt.imgsz *= 2 if len(opt.imgsz) == 1 else 1  # expand
    print_args(vars(opt))
    return opt


def main(opt):
    check_requirements(exclude=('tensorboard', 'thop'))
    run(**vars(opt))


if __name__ == "__main__":
    opt = parse_opt()
    main(opt)
```

Appendix C

Gantt Chart

Gantt Chart

To assure the success of the project one of the significant parts is planning what needs to be done and setting an appropriate time for each task. With this, the timeline of the research study is shown for better management of the study.

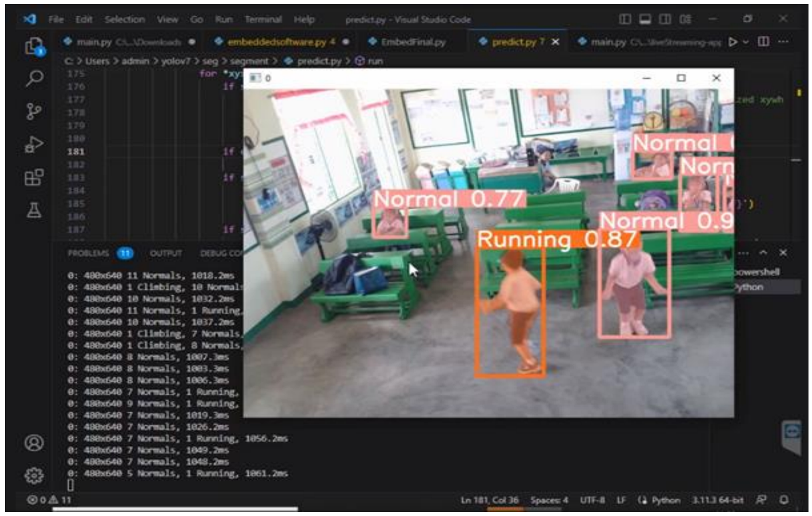
Table 11: Research Timeline

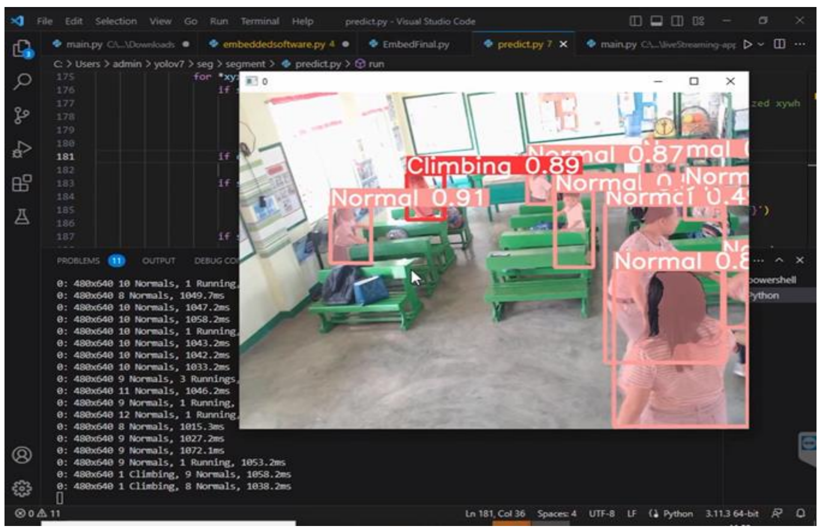
PROJECT GANTT CHART																								
Task	January (2023)				February (2023)				March (2023)				April (2023)				May (2023)				June (2023)			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Equipment Procurement	■	■	■																					
Beneficiary Permission Acquisition				■																				
Image Dataset Collection				■	■																			
Dataset Annotation						■	■	■																
Python Code Construction						■	■	■																
Model Development									■	■	■	■												
Hardware Integration													■	■	■									
Live Testing																■	■	■						
Device Finalization																			■	■	■			
Project Transfer																						■		

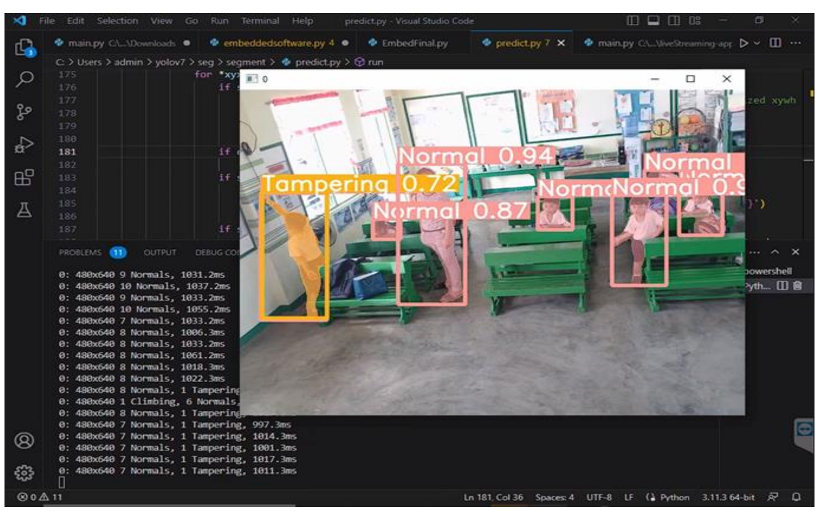
The Gantt chart above shows the timeline of the project from beginning to end. It includes all the processes in order and each step to complete the project. The different stages of the activities are Planning, Data Gathering, Analysis of Data Gathering, Proposal Defense, Implementation Phase, Final Testing, Finalizing of Manuscript, and Project defense. This would help the researchers in planning, managing, and what task is required to complete the project.

Appendix D

Validation


predict.py - Visual Studio Code
Normal 0.77
Running 0.87
Normal


predict.py - Visual Studio Code
Climbing 0.89
Normal 0.91
Normal


predict.py - Visual Studio Code
Normal 0.94
Tampering 0.72
Normal 0.87
Normal

Appendix E

Curriculum Vitae

LOURYDEM DAWIS

COMPUTER ENGINEERING STUDENT

09266708337

lourydem.dawis@g.batstate-u.edu.ph

San Teodoro, Oriental Mindoro

https://www.linkedin.com/in/lourydem-dawis

PROFILE

Highly skilled Computer Engineer with a strong background in programming, machine learning, and project management. Proficient in various programming languages and experienced in developing innovative solutions. My project management skills allow me to effectively plan, organize, and execute complex projects, ensuring successful outcomes. I aim to utilize my skills in a fast-paced industry, driving technological progress and improving user experiences.

SKILLS

- Java
- Python
- HTML
- CSS
- Javascript
- AutoCAD
- Basic CISCO
- Figma
- Deep Learning
- Machine Learning

SEMINARS

- **Robotics Process Automation**
  ICpEP - Singapore Chapter
- **How to Adapt DevSecOps Successfully**
  ICpEP - Singapore Chapter
- **Introduction to Lavalust: A Lightweight Web Framework Based on MVC Pattern**
  Mindoro State University - Calapan City Campus
- **Cloud Computing 101: Opportunities and Challenges**
  Batangas State University- Alangilan
- **Developing Developers: Igniting Students in Web Development Industry**
  Batangas State University- Alangilan

EDUCATION

SECONDARY SCHOOL

San Teodoro National High School

2013-2019

TERTIARY

Batangas State University – Alangilan

2019-2023

PROJECTS

- **Defect Detection and Classification using mask-RCNN and Machine Vision on Printed Circuit Board** (Mask R-CNN, Python, Deep Learning)
- **Prediction of Mean Parameter: An Applied Regression Analysis Using Ensemble Model** (KNN, Linear Regression, Decision Tree Regressor, Python, Machine Learning)
- **An IoT-Based Alarm and Real-Time Monitoring System** (Yolov7, Python, Deep Learning)

JAN MILLER F. JARO

COMPUTER ENGINEERING STUDENT

09477694424

janmiller.jaro@g.batstate-u.edu.ph

Lipa City, Batangas

https://www.linkedin.com/in/janmillerjaro/

PROFILE

Highly skilled and versatile Computer Engineer with a strong background in graphic design, programming, machine learning, UI/UX design, and Web Development. With a passion for innovation and problem-solving, I bring a unique blend of technical expertise and creative flair to every project I undertake. My goal is to leverage my diverse skill set to develop cutting-edge solutions that optimize the user experience and drive business success.

SKILLS

- Java
- Python
- HTML
- CSS
- Javascript
- Deep Learning
- Adobe Photoshop
- Adobe Lightroom
- Adobe Indesign
- Figma
- Machine Learning

SEMINARS

- **Huawei Seeds for the Future**
  Huawei Philippines
- **Developing Developers: Igniting students in Web Development industry**
  Batangas State University- Alangilan
- **Develop and Operate: Expanding Future Engineers' Skills**
  Batangas State University- Alangilan
- **HOP ON: An Introduction to Cloud Computing, The Pilot Technologies**
  Batangas State University- Alangilan

EDUCATION

SECONDARY SCHOOL

Manuel S. Enverga University Foundation

2012-2019

TERTIARY

Batangas State University – Alangilan

2019-2023

PROJECTS

- **Build A Droid** (Java, Codecademy)
- **Desert Island Playlist** (Java, Codecademy)
- **Bluetooth Robot Car** (Arduino, C++)
- **Laevoda: Laundry Service Management Mobile Application** (Android Studio, Java)
- **AlterEye: Footpath and Obstacles Detector System** (Mask R-CNN, Python, Deep Learning)
- **An IoT-Based Alarm and Real-Time Monitoring System** (Yolov7, Python, Deep Learning)

Purok 9, San Vicente Bauan, Batangas, Philippines

janelyncymazo16@gmail.com

+639159393604

SKILLS

- Good programming skills(Java, Python, Javascript, HTML and C++)
- Can utilize Postman, Jira, and ALM
- Good mathematical skills
- Creative problem solver and logical thinker
- Fast learner
- Good technical skills
- Ability to work under pressure
- Ability to work and collaborate with the team
- Attention to detail
- Can sit for a long period of time
- Excellent time management
- Flexible and punctual
- Committed, hardworking, and with a sense of professionalism

INTEREST

- Programming
- Designing and developing apps or website
- Logic and mathematical aspects
- Youtube videos(Computer related)

PERSONAL INFORMATION

Date of Birth : October 16, 2000 Nationality : Filipino
Gender : Female
Marital Status : Single

CAREER OBJECTIVES

Creative Backend Developer wanting to be a self-motivated IT expert with vast knowledge in JavaScript, HTML, and CSS, as well as great talents and abilities in producing clean and efficient code. Having the desire to understand API, as well as Front-End Development, and applying it effectively for the company's satisfaction.

SEMINARS/TRAININGS/ATTENDED

"INTRODUCTION TO CYBERSECURITY"
MAPUA UNIVERSITY
JUNE 30, 2021

"TECHNOLOGY PATH FOR ASPIRING JAVA DEVELOPERS"
MAPUA UNIVERSITY
JULY 15, 2021

"HUAWEI SEEDS FOR THE FUTURE"
HUAWEI PHILIPPINES
NOVEMBER 15-23, 2021

"DEVELOPING DEVELOPERS: IGNITING STUDENTS IN WEB DEVELOPMENT INDUSTRY"
BATANGAS STATE UNIVERSITY
SEPTEMBER 6, 2022

"DEVELOP AND OPERATE: EXPANDING FUTURE COMPUTER ENGINEERS' SKILLS"
BATANGAS STATE UNIVERSITY
SEPTEMBER 27, 2022

"GUARD UP: AN INTRODUCTION TO CYBERSECURITY AND CLOUD SECURITY"
BATANGAS STATE UNIVERSITY
OCTOBER 27, 2022

"HOP ON: AN INTRODUCTION TO CLOUD COMPUTING, THE PILOT OF TECHNOLOGIES"
BATANGAS STATE UNIVERSITY
DECEMBER 1, 2022

"UNFOLD: A DEVELOPMENTAL FOUNDATION TO DATABASE, WEB API AND INFOR ERP WORKFLOW"
BATANGAS STATE UNIVERSITY
DECEMBER 3, 2022

EDUCATION

BS Computer Engineering
Batangas State University
2019-Present