

Article

Not peer-reviewed version

Drowsiness Detection of Construction Workers: A Proactive Approach to Accident Prevention Leveraging Yolov8 Deep Learning And Computer Vision Techniques

[Adetayo Olugbenga Onososen](#)*, Innocent Musonda, [Damilola Onatayo](#), Abdullahi Babatunde Saka, [Samuel Adeniyi Adekunle](#), [Eniola Onatayo](#)

Posted Date: 10 October 2024

doi: 10.20944/preprints202410.0764.v1

Keywords: Construction; Deep Learning; Drowsiness; Construction safety; Computer vision; Accident; Yolo



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Drowsiness Detection of Construction Workers: A Proactive Approach to Accident Prevention Leveraging Yolov8 Deep Learning And Computer Vision Techniques

Adetayo Olugbenga Onososen ^{1,*}, Innocent Musonda ², Damilola Onatayo ³,
Abdullahi Babatunde Saka ⁴, Samuel Adeniyi Adekunle ⁵ and Eniola Onatayo ⁶

¹ Centre of Applied Research and Innovation in the Built Environment (CARINBE), Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg 2092, South Africa 1

² Centre of Applied Research and Innovation in the Built Environment (CARINBE), Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg 2092, South Africa 2

³ Department of Construction Management, University of Florida

⁴ Westminster Business School, University of Westminster, 35 Marylebone Road London GB, NW1 5LS

⁵ Cidb Centre of Excellence, University of Johannesburg, Johannesburg, South Africa

⁶ Department of Environmental Engineering, State University of New York

* Correspondence: 221046022@student.uj.ac.za

Abstract: Construction projects' unsatisfactory performance has been linked to factors influencing individuals' well-being and mental alertness on projects. Drowsiness is a significant indicator of sleep deprivation and fatigue, being able to identify the cognitive and physical preparedness of workers on site to engage in construction tasks. As a consequence of the strenuous nature of the work involved in construction, long work hours, and environmental conditions, drowsiness is commonplace and has received less attention despite being a leading cause of accidents occurring on-site. Detecting drowsiness is essential for determining the safety and well-being of site workers. This study presents a vision-based approach using an improved version of the You Only Look Once (YOLOv8) algorithm for real-time drowsiness exposure among construction workers. The proposed method leverages computer vision techniques to analyse facial and eye features, enabling early detection of drowsiness signs, effectively preventing accidents, and enhancing on-site safety. The model showed significant precision and efficiency in detecting drowsiness from the given dataset, accomplishing a drowsiness class with a mean average precision (mAP) of 92%. However, it also exhibited difficulties handling imbalanced classes, particularly the underrepresented 'Awake with PPE' class, which was detected with high precision but comparatively lower recall and mAP. This highlighted the necessity of balanced datasets for optimal deep learning performance. The YOLOv8 model's average mAP of 78% in drowsiness detection compared favourably with other studies employing different methodologies. The vision-based drowsiness detection system has broad applications in the construction industry. It can be integrated into existing safety protocols, enabling real-time alerts to supervisors or workers when drowsiness is detected. The system improves productivity and reduces costs by preventing accidents and enhancing worker safety. However, limitations, such as sensitivity to lighting conditions and occlusions, must be addressed in future iterations.

Keywords: construction; deep learning; drowsiness; construction safety; computer vision; accident; yolo

1. Introduction

As ensuring safety is paramount in all infrastructure projects, there is a pressing requirement for implementing on-site safety regulations and protocols. This is essential to enhance the well-being and security of construction workers, as highlighted in the works of [1,2]. This can be ascribed to the fact that the construction site is one of the most hazardous environments, with several safety concerns stemming from the intricate and fluid interaction that exists between employees, materials, equipment, and the actual execution of construction tasks [3,4]. Fatigue stands out as a prominent contributor that jeopardises the health, safety, and overall welfare of individuals working within the construction industry. Fatigue, which results from lack of sleep, negatively impacts a person's well-being, ability to function at work, and safety [5,6]. As indicated by prior research conducted by the National Safety Council, every single surveyed construction worker exhibited at least one susceptibility to workplace fatigue, a condition that can result in hazardous work environments and heightened injury hazards [5]. 71% of construction industry employers who responded to the survey claimed that their employees' lack of sleep had an effect on productivity, and 45% of respondents said that weariness was to blame for safety-related incidents [5,6].

Drowsiness is an indication of fatigue, which can have a substantial negative impact on a person's well-being, productivity, and ability to stay safe. Fatigue is a difficult ergonomic/safety "issue" in several sectors such as manufacturing [7], Construction [8,9], and others [9] since it reduces productivity and raises the likelihood of accidents. These consequences could result in decreased performance, decreased production, deficiencies in the quality of the work, and human errors.

Construction tasks often entail heavy workloads, uncomfortable working postures, and long working hours, making construction employees susceptible to weariness and drowsiness [8,9]. Draws the significance of drowsiness research by highlighting the long-term consequence of unchecked fatigue as leading to chronic fatigue symptoms and reduced immune abilities. It has also been pointed out that these outcomes have been highlighted with reduced quality of life, socioeconomic disruption to way of life, increased absenteeism etc. However, while drowsiness holds significant implications for the well-being of workers and project performance, its detection among construction workers is low [10]. Traditional methods of on-site sleepiness assessment to combat fatigue typically rely on a visual examination carried out by a trained observer. This is despite the fact that drowsiness is becoming an increasingly significant topic in the fields of health, safety, and well-being research [11]. These approaches do not effectively grasp the interconnectedness of numerous risk elements and the diversity in the tasks undertaken.

Despite its severity, numerous serious workplace mishaps continued to have been linked to insufficient sleep. Still, a common challenge is that people often do not comprehend their level of fatigue, its effects, or both. Previous studies such as those [3] have discovered that adults need 8 hours of sleep daily to be fully restorative. Yet, most only receive 7 hours, leaving them with a sleep deficit. Therefore, it is imperative to develop innovative methods to identify drowsiness on site [12]. This is critical to reducing the likelihood of accidents on site as drowsiness impairs attention, reaction time and judgement. The need for innovative ways to identify the fatigue state through drowsiness has further been postulated to be vital to avert; falling from heights [13], slow reflexes leading to a collision with equipment [14], workers inability to operate machinery [15], poor decision making [16], inefficient work performance [9], and elevated stress levels amongst others [17].

Therefore, observation of risky situations and action is necessary for construction safety and health to eliminate possible risks quickly [18]. However, such observations become ineffective and laborious when using subjective and manual assessment methods. Such manual assessments of the worker's safety conditions are laborious and error-prone since they depend so much on the inspector's physical condition and level of knowledge [19]. This has required the development of inventive digital methods that make use of computer vision and deep learning.

In recent times, computer vision methods have emerged as a reliable and automated means of conducting field observations, extracting safety-related insights from images and videos recorded on the site. These approaches are seen as useful replacements for manual observational methods, which take a lot of time and are not very trustworthy [16,19]. The current surge in interest and uptake of computer vision can be attributed to its capability for automated and continuous monitoring of

construction sites. By taking pictures or videos of a construction scene, computer vision may reveal a wealth of information about it, such as the locations and behaviours of project entities and the circumstances of the site [20]. This has the potential to enable a quicker, more precise, and all-encompassing comprehension of complex construction activities [13]. As a consequence of this, computer vision has been integrated into a number of distinct parts of the construction industry, such as the tracking of progress, the evaluation of productivity, the identification of defects, and the automation of documentation [19,21]

With the aim of overcoming the constraints associated with traditional manual methods for detecting drowsiness as a fatigue indicator, this research employed computer vision and deep learning techniques. This enabled the early identification of signs of drowsiness, offering a potent approach to accident prevention and the enhancement of on-site safety. This study contributes to improved health, safety, and well-being on construction sites by developing a vision-based approach using an improved version of the You Only Look Once (YOLOv8) algorithm for real-time drowsiness detection in construction workers. The approach was based on the preparation of datasets, the training of deep learning models, and the evaluation of the models using testing images. Specifically, the awakesness pose estimate dataset is acquired from photographs and videos recorded on actual and non-actual construction building sites. Subsequently, critical key points encompassing complete equipment body postures are delineated and labelled within the collected images. The results of the experiments demonstrated that the suggested methodology framework can rapidly and accurately anticipate the whole drowsiness postures of construction workers. The model displayed significant precision and efficiency in detecting drowsiness from the provided dataset. It achieved a mean average precision (mAP) of 92% and an inference speed as low as 0.4ms for preprocessing and 7.5ms for inference. Both of these figures are quite low. However, it also exhibited difficulties in handling imbalanced classes, particularly the underrepresented 'Awake with PPE' class, which was detected with high precision but comparatively lower recall and mAP. This highlighted the necessity of balanced datasets and appropriate hyperparameters for optimal deep-learning performance. The YOLOv8 model's mAP of 78% in drowsiness detection compared favourably with other studies employing different methodologies.

The subsequent sections of the document are organized as follows: Section 2 delves into a thorough review and discussion of previous research; Section 3 offers an extensive depiction of the proposed methodology; Section 4 showcases the outcomes achieved using deep learning techniques. The fifth and concluding section of the paper presents a summary of the research's core content, underscores its implications, addresses limitations, and outlines potential avenues for future exploration.

2. Construction Workers Well Being

The construction industry exerts a profoundly influential effect on the economic prosperity of every country, as demonstrated by its impressive \$10 trillion addition to the global gross domestic product (GDP), as reported by [19]. Nevertheless, because of the large number of accidents and fatalities that occur within the industry, its influence is negatively impacted as a direct result. This calls for an urgent need to improve its workforce's health, safety, and well-being operations. A worker's performance suffers due to fatigue in the workplace, which is a multifaceted construct [23]. It is the product of continued action and is influenced by psychological, socioeconomic, and environmental factors [6]. Recent evidence suggests that improved operational practices that have enhanced workers' wellness and safety are influenced by studies on improving the well-being of workers [15,23,24].

Despite the fact that working on a construction site while feeling overly fatigued could be risky, just 75% of construction employees believed that it was, but 98% of construction employers concurred [5]. In addition, there was a disparity between the percentage of employees (78%), who believed that driving when fatigued is risky and the percentage of employers (96%), who agreed with this sentiment. Most construction employees (76%), who identified job demands as a risk factor for weariness, indicated they were affected by them. The number two spot was taken by lengthy

commutes, with 46% of workers considering this a contributing factor. Working during nighttime or in the early morning hours ranked third at 46%, following factors such as receiving less than seven to nine hours of sleep per night (41%), working 50 or more hours per week (28%), and working shifts lasting 10 hours or more (27%). This was followed by obtaining less than seven to nine hours of sleep per night. According to the findings of a report compiled by the National Safety Council on “Fatigue in Safety-Critical Industries — Impacts, Risks & Recommendations” [5]. The challenge this demonstrates is the disparity between workers’ responses in subjective assessment of their state of fatigue and actual readiness to engage in construction tasks. Recent evidence by FinancialTimes, [25] revealed that between 400 and 500 workers died due to the 2022 world cup projects. This indicates how grave infrastructure delivery can turn out when the fatigue assessment is only left to subjective measures.

2.1. Unsafe Behavior in the Construction Sector

As per the findings of Yu et al. [8], about eighty percent of construction accidents are the result of risky practices carried out by workers. This can include unsafe behaviour on the part of individuals, unsafe behaviour on the part of teams, and unsafe practices in relation to machines, equipment, and robots [1,2] Due to the restrictions associated with traditional methodologies, it has been difficult for project and construction managers to monitor workers’ unsafe behaviours; as a result, innovative applications are being launched to increase the efficiency and efficacy of unsafe behaviour monitoring. This involves the development of technologies such as motion capture technologies [8], computer vision and deep learning [26], and wearable robots [13]. As stated by Yu et al. [8], some examples of unsafe behaviours on construction sites are; sleeping on sills and pulling trolleys on stairs. Others include drowsiness, Working at heights without proper fall protection, improper use of tools and equipment, disregarding safety guidelines and procedures, etc. However, most incidents can be reduced with increased alignment and obedience to required safety principles and guidelines [27]. Although governing laws and safety regulations, such as OSHA-1926.28a, typically place the responsibility on employers to enforce, monitor, and uphold proper safety protocols, as highlighted by Nath et al. [27], employees often neglect to adhere to these regulations on the job site. This disregard can stem from a lack of awareness regarding safety measures, discomfort associated with wearing personal protective equipment (PPE), and the belief that PPE hampers their job performance. Vision-based and sensor-based automated PPE compliance monitoring technologies are the two primary types of this type of monitoring technology now available. In approaches that are based on sensors, a sensor will be installed, and the signals it produces will be analysed. As an example, one approach could involve affixing RFID tags to individual pieces of personal protective equipment (PPE), and positioning a scanner at the workplace’s entrance to read these tags. This would help ascertain whether employees are complying with the mandatory PPE requirements.

The ever-changing and intricate characteristics of construction projects almost certainly increase workers’ dangers on the job. Without systematic and comprehensive safety and health management procedures in place on construction worksites, it is unfeasible to completely eradicate occupational risks, as emphasized by Seo et al. [18]. These measures encompass safety planning, worksite analysis, prevention and control of hazards, as well as safety and health training. Consequently, it becomes necessary to implement actions like monitoring unsafe practices. Monitoring risky situations and behaviours during construction is crucial to identifying them and acting quickly to avert further safety and health problems by removing them from the causal chain [9]. In the practice of construction, site observations and inspections are frequently utilised in order to determine the level of risk that is linked with ongoing projects and the present state of the site. Observational methods are costly and time-consuming due to the fact that they require supervisors or safety employees to manually make and record observations. According to Seo et al. [18], one of the limitations of manual observation is that it cannot provide timely access to information that is either incomplete or incorrect.

2.2. Fatigue amongst Construction Workers

The majority of the job that site employees perform in the construction industry is repetitious and taxing physically. According to Ray and Teizer [9], people who perform this kind of job in abnormal postures put unnecessary strain on their body parts, which can lead to weariness and drowsiness, both of which can result in injuries or, in the most extreme situations, permanent disability. In accordance with [22], fatigue is characterized as a shift in task performance resulting from the initial mental and/or physical exertion that is so demanding on the worker's comfort that it hinders their ability to meet the demands placed on their cognitive functioning. As previously mentioned, occupational fatigue has a variety of components, including both physical and mental fatigue.

Physical fatigue is a decrease in one's capacity to do a physical task due to earlier physical activity [6,24,25]. Fatigue may be especially problematic in the built environment since it reveals discomfort, impaired motor function, and decreased strength capability. After completing physical duties for an extended amount of time, a person becomes physically exhausted, which eventually affects their capacity to conduct physical tasks successfully [23]. Also, doing mental tasks for a prolonged time causes mental tiredness, which reduces a person's cognitive capacity. Conversely, Ibrahim et al., [23] argue that there are limitations to different means of measuring fatigue. Anwer, Li et al., [21] presents an account of fatigue measurement using subjective and physiological metrics. However, it is difficult to assess fatigue on the job site due to the building site's dynamic nature and the wearable sensors' intrusive nature. Contemporary approaches to detect and track physical fatigue typically hinge on invasive monitoring of brain activity, such as employing electroencephalography (EEG), or recording sleep habit diaries to evaluate whether the worker possesses the requisite capacity prior to commencing their tasks. When workers are fatigued, their ability to accurately comprehend job-related information that could potentially endanger them becomes more challenging. This heightened difficulty in comprehension elevates the risk of accidents occurring in that specific work environment. According to [22], there have been very few occupational applications directly related to the detection of physical exhaustion in the most physically demanding occupations. Some examples of these occupations are construction, manufacturing, and agriculture.

2.3. Drowsiness by Workers in the Construction Industry

Sleep deprivation can be brought on by a wide variety of circumstances, such as the choices one makes in their lifestyle, the effects of stress, bad sleeping habits, and sleep disorders like sleep apnea and restless legs syndrome. Whatever the cause may be, sleep deprivation and a lack of sleep can have a negative impact on performance, which in turn raises the risk of accidents for both the worker and others in the workplace [3]. The major consequence of drowsiness on construction sites is its likelihood of leading to accidents or wrong decision-making, affecting project performance and the risk of chronic health challenges [9].

According to Seo et al. [18], carrying out job safety observations and inspections is one of the most popular methods that is utilised in the construction industry to evaluate ongoing operations. These actions are included in a more comprehensive safety and health monitoring category. During the observation, the human observer serves to detect and eliminate the potential causes of accidents (i.e., unsafe conditions and acts) by watching workers perform a specific task (i.e., safety observation) or visually inspecting the work area and work equipment (i.e., safety inspection) with a checklist [12]. This is accomplished by observing workers as they carry out a particular task (also known as "safety observation") or by visually inspecting the workspace and the equipment that is used for the job (also known as safety inspection) [8].

Questions have been raised about the quality of sleep needed to remove drowsiness; previous studies have postulated that adults need 8 hours daily to be fully restorative. Yet, most only receive 7 hours, leaving them with a sleep deficit [22]. Sleep deprivation occurs when an individual gets less sleep than is necessary for complete recovery. Serious or ongoing sleep deprivation can harm the person experiencing it and anyone else affected by their behaviour. While increased attention is paid to task-specific and known risks, it has been identified as a significant reason why the number of

fatalities that occur in the construction industry keeps rising. One example of this is the failure to take into account unknown or conspicuous dangers, such as drowsiness [23]. Effective elimination of safety and well-being risks on sites depends on preventive strategies for easily identified and known risks and not easily perceived or recognised risks such as drowsiness [27–29].

Studies done in the past have also investigated the connection between a lack of situational awareness and tiredness while on the job. Drowsiness in workers can have a significant impact on their situational awareness, which can cause difficulties for workers in recognising, perceiving, and analysing hazards, as well as in making projections and establishing control [30]. Therefore, effective identification of indicators of fatigue, such as drowsiness, is imperative. Ibrahim et al. [23] identified recognising hazards as the awareness that a situation might be dangerous, with two forms of hazard recognition being; predictive and retrospective. This study takes a predictive approach to fatigue recognition utilising drowsiness as an indicator. It aims to build a vision-based strategy employing an upgraded version of the You Only Look Once (YOLOv8) algorithm for real-time drowsiness detection in construction workers. The retrospective technique analyses data from past safety incidents to in preventing future recurrences, whereas the predictive approach forecasts future working scenarios and predicts safety threats.

Previous development in this area has revealed that by leveraging deep learning algorithms, mental alertness due to drowsiness can be effectively measured compared to other less effective methods, such as physiological metrics. This has been demonstrated by [31], who utilised pressure sensors to measure subject fatigue, and a combination of deep learning algorithms and biomechanical analysis was employed to provide a non-intrusive method of monitoring the physical exhaustion of the complete body while the construction process was taking place. Other methods of determining mental exhaustion entail constructing a model that can objectively identify the level of mental weariness in construction workers based on data from the wearable electroencephalogram sensors that were administered to 15 participants [32]. It is becoming increasingly common to apply computer vision techniques in safety monitoring in addition to sensor-based methods [12,33].

A multi-sensor data monitoring system is required because of the complex nature of the factors contributing to fatigue development. In order for technological approaches to the measurement of physical fatigue to be effective, the system must be able to predict physical fatigue (before it has a negative impact on productivity or safety), measure and monitor physical fatigue in the operational environment, and enable intervention when deficits are discovered or foreseen with the use of appropriate interventions [34]. According to Mariam et al. [11], one of the obvious measures for detecting physical weariness in the workplace is to ask the worker to rate their perceived level of physical fatigue. However, the practice has shown that the majority of workers report incorrectly in order to avoid being replaced on duties. Hence, self-reported fatigue is largely limiting in predicting construction workers' mental alertness and cognition.

Drowsiness is imperative to be measured amongst construction workers for a couple of reasons; due to hazardous weather, use of mobile equipment, several work sites needing travel, and unpredictable and demanding schedules mandating extra work hours, risks on construction sites may be increased [9]. All of these might heighten fatigue and exhaustion. These factors strain human physiology and can cause weariness that impairs performance and increases worker risk [11]. When performance and decision-making are impaired, this not only puts the worker at safety risk but places the tasks being conducted at risk, which can have a monumental effect on the structure's integrity. While previous studies have given valuable contributions to the impact of fatigue on workers' operations, little is still known about identifying drowsiness as an early indicator of mental and physical fatigue.

2.4. Computer Vision Techniques and Deep Learning in Construction

The field of computer vision is one that draws from several disciplines that investigates the ways in which computers can gain significant knowledge from viewing digital images or videos. Deep learning approaches have recently attracted a lot of interest in the field of computer vision due to their capacity to acquire valuable features on their own from enormous volumes of annotated

training data [27]. This ability has contributed to the increased popularity of deep learning techniques. The reason for this is that approaches of deep learning have the capability of improving themselves by studying their past errors. It intends to automate tasks that, from an engineering point of view, the human visual system is unable to execute [35,36]. According to Seo et al. [18], computer vision-based safety and health monitoring require the capture of photos or videos of the construction sites where the activity to be monitored is taking place. This is a prerequisite for the monitoring process. For the purpose of getting the 2D imaging data (sometimes referred to as 2D videos or sequential images), which is important for computer vision-based monitoring, the photos or videos may serve as low-cost alternatives.

Considered to be one of the most essential components of computer vision is the criteria for recognising object tracking on building sites (Seo et al., 2015; Akinosho et al., 2020). Several parameters that need to be considered include frame rate, outdoor application capabilities, reliable reading range, object localization capability, and 3D modelling capacity. Second, in order to verify the identities of the workers and determine whether or not they are licenced to carry out the work, we utilise facial detection and recognition methods. These procedures help us determine whether or not the workers are authorised to carry out the task. CNN enables the automatic recognition of a large variety of objects contained inside an image, which is a vital step for ongoing study. These stages essentially consist of face detection and recognition, object detection and tracking, as well as object detection and tracking [23,37]. Face detection is required before moving on to the next step of face recognition. According to Cha et al. [31], the classification of facial recognition performed by computers requires images that are up close and solely show the face. Computer vision approaches are more versatile and adaptable than sensors because they do not require workers to wear additional equipment [38,39].

This was made much simpler by the development of new algorithms such as Faster R-CNN, which are able to recognise and keep track of resources such as people, plants, and equipment, as well as identify personnel who are behaving in a risky manner [39–41]. According to Fang et al. [12], action recognition is the most important aspect of computer vision-based systems. These systems make use of manually constructed features (such as shapes) in images or videos. Image representation systems that are used to recognise human behaviours are able to extract information from images, such as shapes and temporal motions, which are used to do so. In order to correctly identify and evaluate a wide range of actions, the action identification features must contain extensive information. Classifier tools (such as Support Vector Machines [SVM]), temporal state-space models (such as Hidden Markov models [HMM] and conditional random fields [CRF]), and detection-based techniques (such as bag-of-words coding) can all be used to assess such properties [12]. Due to the advantageous trade-off between model correctness and inference speed, the You Only Look Once (YOLO) series algorithm models 30–32 are recommended for use in real-time applications [13,40].

Edge detection on images was one of the key methods that were utilised for excavator pose estimates in the early days [15] Also, [19] employed silhouette-based tracking algorithms to extract binary images from videos captured by stationary security cameras to estimate trolley movement along the crane jib. These algorithms relied on the films captured by the cameras. [20,41] employed a non-rigid equipment posture estimation based on construction pictures and videos to establish a model for detecting equipment parts using support vector machines (SVMs) and histogram-oriented gradient (HOG). This model was developed for detecting equipment parts using support vector machines (SVMs) and histogram-oriented gradient (HOG). By incorporating the k-means methodology, background subtraction algorithm, and the part-based posture estimation method, Soltani et al., [30] improved it and produced more accurate findings. In another study conducted by Soltani et al. [30,41], the time and coordinate systems of multiple cameras and the real-time location system (RTLS) were synchronised. This allowed the RTLS to combine the data from those sources, which allowed the researchers to extract two-dimensional equipment poses and estimate the three-dimensional poses of excavators.

Deep learning strategies have been increasingly popular for use in a variety of occupations within the construction sector. These jobs include the monitoring of construction sites and the health

inspection of civil infrastructures. These jobs were created to address challenges that were previously resolved using normal computer vision methods. Convolutional neural networks, also known as CNNs, are a type of neural network that is frequently utilised in deep learning approaches [33,42]. According to Fang et al. [12], approaches for deep learning that are based on CNN are effective for computer vision and pattern recognition. Concrete crack detection using CNN-based approaches has been used for civil infrastructure health inspection; the results showed that CNN is superior to other computer vision methods [19,29].

An algorithm for deep learning was developed with the help of a faster Region-Based Convolutional Neural Network (Faster-RCNN), which was used to identify four distinct types of sewer pipe failures. This highlights the benefits of using deep learning algorithms for analysing images and videos in even more detail. LeNet-5 is a CNN model that was constructed by LeCun and his colleagues that deciphers handwritten digits. The dataset that was used to build the model was produced by the mixed National Institute of Standards and Technology. CNN models are able to efficiently and automatically identify attributes from static images (Fang et al., 2018; Akinosho et al., 2020). This capability is achieved by stacking several convolutional and pooling layers.

Fang et al. (2018) created two algorithms based on Faster-R-CNN and CNN models that are used to (1) recognise the presence of employees and (2) decide the harness that is fastened to them in an effort to solve the problem of workers working at heights forgetting to wear their harnesses. This was done in order to address the problem of workers working at heights and forgetting to wear their harnesses. Similar to this, Fang et al. [12] developed a deep learning method to automate the process of inspecting the use of personal protective equipment (PPE) by stepladders in aerial works. Nath et al. [27] constructed three deep learning (DL) models built on You-just-Look-Once (YOLO) architecture to verify the PPE compliance of employees and [29] completed an automated examination of large-scale bridge constructions just using photos.

Other applications of deep learning include the application of two CNN models with an extremely high level of accuracy to recognise safe harnesses worn by workers to prevent falling from heights [12]), the recognition of unsafe behaviours [42], the estimation of the poses and activities of construction employees [19,43], and the recognition and tracking of equipment [26]. On the other hand, CNN has served as the foundation for virtually all of the effective algorithms that have been created for the purpose of image categorisation, object detection, and visual tracking. Because of this, their use in the construction industry for use cases involving visual detection has been further promoted. The findings of these studies offer valuable insights into improving building processes through the application of deep learning and computer vision techniques. Even though the relevant research is still in its early stages, the current state of the art implies that deep learning and computer vision techniques offer substantial potential for monitoring even fatigue-associated instances such as drowsiness. This is despite the fact that the relevant research is still in its early stages.

3. Methodology

In this study, we elaborate on developing and implementing a vision-based model that is adeptly trained on a substantial corpus of visual data, encompassing both videos and still images. This methodology has been widely used in similar studies that were conducted in the construction industry [29,34]. This information was gathered from potential construction operatives, as well as numerous construction operatives, all within the framework of the South African construction sector. To engineer a rich and varied dataset, video footage acquired from the above sources was systematically processed to generate additional image data, thus enhancing the density of our dataset. Notably, the dataset includes 13 distinct operatives, each meticulously recorded to encapsulate a range of activities. These activities were primarily aimed at simulating indicators of drowsiness, thereby promoting a model attuned to early recognition of potential safety hazards.

In this vein, the operatives were requested to perform and simulate various activities indicative of drowsiness. These include, but are not limited to, yawning, blinking at a slower rate, lethargic head movements, and showing visible signs of fatigue in the eyes. Each of these activities is a well-documented precursor of drowsiness, a state of potential risk in the challenging working conditions

of the construction industry [33]. The recording was executed using high-resolution cameras that served dual purposes to capture the richest possible data. They were employed to acquire still images and to document videos concurrently. The quality of the recorded data was maintained at an impressive level, with the video content encapsulating approximately 1.5 hours of footage, all recorded at a high resolution of 1080 x 1920. Furthermore, this data was captured at a rate of 28 frames per second, ensuring clear, detailed visual content. A comprehensive flowchart of the model training procedure, validations, and data set testing are presented in Figure 1. This is done in order to accomplish the goals of the study.

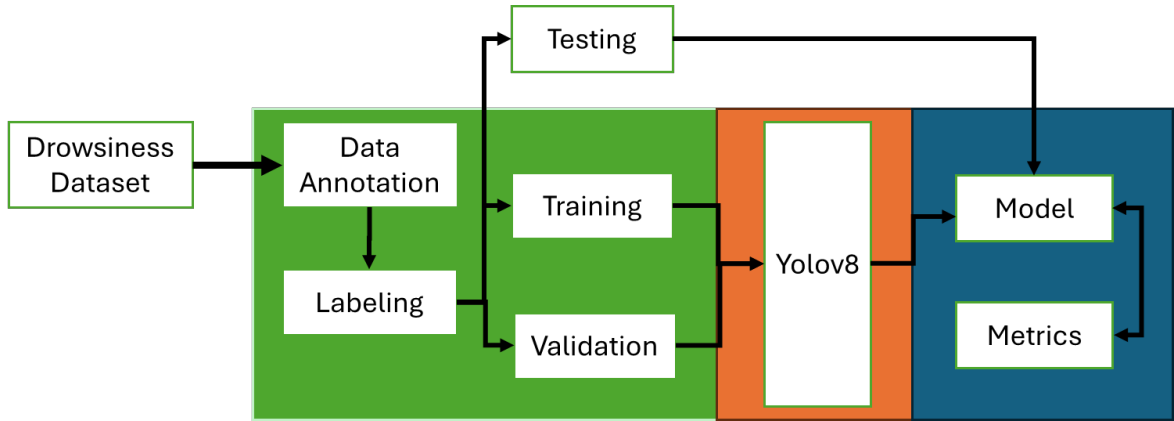


Figure 1. Flowchart of the model training, validation and testing on the dataset.

3.1. Dataset Extraction

The process of dataset extraction involved the meticulous extraction of still images from the recorded videos. These images were subsequently employed as inputs to the Yolov8 model. We set the interval between the extracted frames at 36 seconds to maintain consistency and ensure an evenly spaced data distribution. This interval was chosen after a series of experimental observations, aiming to maximise the quality and representativeness of the data whilst minimising redundancy, as shown in [34,35]. Utilising the robust OpenCV algorithm, these frames were converted into still images, yielding approximately 149 images. In addition to this, we also incorporated self-captured images, thereby increasing the depth and variety of our dataset, resulting in a total of 605 distinct images.

Given the variance in image resolutions within our drowsiness image dataset, we undertook a rescaling operation to standardise the resolution of all images to 416 x 416 pixels. The need to accommodate the pixel resolution requirements of the You Only Look Once framework, a real-time object identification system that was applied in this study, was the primary factor that led to the formation of this conclusion.

Subsequently, the corpus of 605 images was classified into four distinct categories: ‘Awake with PPE’, ‘Awake without PPE’, ‘Drowsy with PPE’, and ‘Drowsy without PPE’. Table 1 shows the distribution of the images.

Table 1. Distribution of the Dataset.

Class	Numbers included
Awake with PPE	98
Awake without PPE	121
Drowsy with PPE	159
Drowsy without PPE	227

In the course of our experiment, we utilised a random data partitioning approach, in which we allocated seventy percent of the dataset to training, fifteen percent to validation, and fifteen percent to testing. The purpose of this split is to evaluate the model’s performance in a way that is both reliable and preventative of overfitting.

3.2. Computation Specifications

The construction operatives' drowsiness detection framework was built, leveraging hardware and software resources. We utilised Tensorflow 2.0, Keras, and OpenCV libraries, owing to their capabilities in implementing and optimising deep learning architectures. Python 3.9, a dynamic and high-level programming language that is well-known for its readability and convenience of use in scientific computing, was used in conjunction with these libraries. Our model architectures, namely the YOLOv5 and CNN (Convolutional Neural Network) models, were trained without using high-performance Graphics Processing Units (GPUs). Rather, we relied on an 11th Gen Intel(R) Core(TM) i7-11800H @ 2.30GHz 2.30 GHz computer, equipped with 16GB of RAM. The Jupyter Notebook was the development environment of choice during model training and evaluation. All the data for our study were collected using a 16MP camera, further attesting to our commitment to maintaining high-quality data.

3.3. Evaluation Metrics

In assessing the performance of our machine learning model, we have elected to employ a set of computational metrics designed to offer nuanced insights into the efficacy of the model. These metrics are particularly relevant in object detection and classification and are grounded in the calculation formulas derived from the datasets at our disposal. The maximum number of batches, denoted as 'Max batches', is calculated as the product of the number of classes and a factor of 2000. Moreover, the steps are determined by a range between 80% and 90% of the maximum batches. The filter parameter is determined by multiplying the value obtained from adding the number of classes to 5 by a factor of 3.

Max batches = number of classes * 2000

Steps = (Max batches * 0.8, Max batches * 0.9)

Filters = (number of classes + 5) * 3

These calculations inform the analysis of the model's performance in classifying objects. The model's effectiveness in correctly predicting a positive class is denoted as a True Positive (TP). In contrast, its ability to correctly predict a negative class is signified as a True Negative (TN). Conversely, an incorrect prediction of a positive class is termed a False Positive (FP), and an incorrect prediction of a negative class is regarded as a False Negative (FN). These are fundamental metrics retrieved from the output of the object detection algorithm [36].

Precision, the ratio of True Positives to the sum of True Positives and False Positives, measures the model's capacity to correctly predict positive instances Aich et al., [37]. Conversely, recall is computed as the ratio of True Positives to the sum of True Positives and False Negatives, offering a measure of the model's aptitude in accurately identifying all positive instances. Mathematically, these are represented according to equations 1 and 2 [38].

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

A further measure of model performance, the Average Precision (AP), is a function of the relationship between precision and recall rates within a given data sample. This measure incorporates the Intersection over Union and assesses the model's ability to identify an object correctly. The average precision (AP) is calculated as the mean of recall rates with precision ranging from 0 to 1, as shown in equation 3.

$$AP = N \sum_{k=1} P(k) \Delta r(k)$$

In this context, "N" denotes the total number of images included in the dataset that was utilised for the calculation., $P(k)$ signifies the precision rate for image k, and $\Delta r(k)$ denotes the difference in recall rate from image (k-1) to image k. The AP is then calculated for each class and averaged to yield the Mean Average Precision (mAP) for all classes combined.

Accuracy, which describes the frequency with which the model correctly classifies a data point, is computed as the ratio of the sum of True Positives and True Negatives to the total count of True

Positives, True Negatives, False Positives, and False Negatives (Aich et al., 2019). Symbolically, this is expressed as follows in equation 4:

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

However, it is important to note that accuracy has limited utility as a performance metric in object detection due to the negligible relevance of True Negatives. Instead, we focus our evaluation on recall and precision measures and construct a Precision-Recall curve to visualise the trade-off between these two metrics [40]. A model exhibiting high recall and low precision suggests a high volume of detections, most of which are incorrectly labelled. Conversely, a model characterised by high precision and low recall indicates a lower volume of detections, most of which are correctly labelled.

3.4. YOLOv8 Model

The YOLOv8 model, a renowned member of the YOLO model lineage, has earned its reputation through its profound capabilities for joint detection and segmentation (Nath et al., 2020; Onatayo et al., 2023b). It shares its architectural construct with its predecessor, the YOLOv7 model, encompassing distinct components such as a backbone, head, and neck (Aboah et al., 2023). However, the YOLOv8 model distinguishes itself with its novel architecture, fortified convolutional layers that comprise its backbone, and a significantly enhanced detection head. These advancements render it an optimal choice for real-time object detection tasks. Complementing this, the YOLOv8 model also extends support for state-of-the-art computer vision algorithms, particularly instance segmentation. This attribute enables the model to detect multiple objects within a single image or video proficiently. The model operationalises the Darknet-53 backbone network, a notable improvement over the network used in YOLOv7 in terms of speed and precision (Wang et al., 2019a). A distinct feature of YOLOv8 is its adoption of an anchor-free detection head for predicting bounding boxes, further enriching its detection capabilities [42,44]

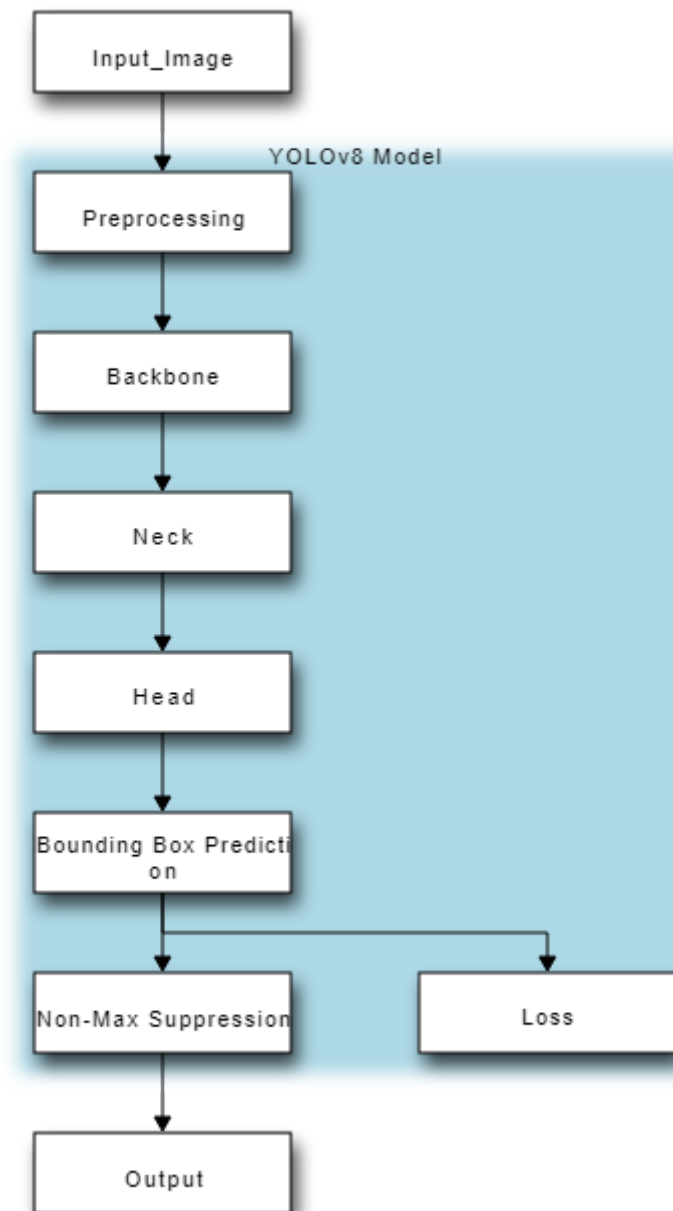


Figure 2. Typical architecture of YOLOv8 algorithm model.

3.5. Dataset Labeling

Two file extensions—.jpeg for images and.txt for text—are used by the YOLO family to identify objects. A text file is utilized to keep track of the labels, object types, and coordinates of their bounding boxes; in contrast, the picture file merely contains images. The image object count is proportional to the row count in the text file. The potential for manually annotating drowsiness on the collected images is labour-intensive and time-consuming. As such, we harnessed the capabilities of LabelImg, an interactive graphical image annotation tool [40]. The labelling tool was instrumental in easing the dataset creation process, allowing for the seamless importation of a series of images. This was followed by the manual delineation of bounding boxes around objects of interest within each image. The identified objects were then classified according to a predefined list of classes specifically curated for this research: 'Drowsiness with PPE', 'Drowsiness without PPE', 'Awake with PPE', and 'Awake

without PPE’. The repetitive process of drawing bounding boxes and assigning class labels was undertaken for all objects within each dataset image. Upon completion, the Labelling tool facilitated the exportation of annotations as text files in the YOLO format. These files contained critical information, including the coordinates of each bounding box and the corresponding label attributed to the enclosed object. Such annotated data serves as an invaluable ground truth for training the detection algorithms, empowering them to identify and classify details within novel, unseen images accurately.

4. Results

The principal aim of the experiment conducted herein was the detection of construction operatives’ drowsiness with high precision and, in real-time, leveraging the YOLO-v8 model to train our datasets. This section elucidates the training process, the variety of hyperparameters employed, the optimisation methods adopted, and the results of the model’s performance. The mode is currently set to “train,” and the job is configured to “detect,” suggesting that the model is currently being trained to recognise objects. The model used is ‘yolov8s.pt’, sourced from ‘data. yaml’. The model is trained for 600 epochs with a patience of 50, indicating the number of epochs with no improvement, after which training will be stopped. The batch size is 16, and the image size is 416. The ‘save’ parameter is set to True, allowing the model to be saved after training. The ‘cache’ parameter is False, indicating that the data is not cached during training. The ‘device’ parameter is set to None, meaning the model will be trained on the default device. The number of worker threads for data loading is set to 2 to ensure the system doesn’t use all the memory.

The optimiser used for training is Stochastic Gradient Descent (SGD). The ‘verbose’ parameter is set to True, enabling the printing of detailed information during training. The ‘seed’ is set to 0 for reproducibility, and ‘deterministic’ is set to True, ensuring deterministic behaviour from PyTorch, cuDNN, and CUDA. The ‘amp’ parameter is set to True, enabling automatic mixed precision for faster training. The ‘fraction’ parameter is set to 1.0, indicating the image fraction to process during training. The ‘overlap_mask’ parameter is set to True, enabling the overlapping mask for mosaic augmentation. The ‘mask_ratio’ is set to 4, and the ‘dropout’ is set to 0.0, indicating no dropout during training. The learning rate is set to 0.01 (lr0), and the final learning rate (lrf) is also 0.01. The momentum is set at 0.937, and the weight decay to 0.0005. The warmup epochs are set to 3.0, with a warmup momentum of 0.8 and a warmup bias learning rate of 0.1. ‘nc=4’ is entered into the model configuration override, which indicates that the number of classes in the model is changed to 4, as described in the methodology section.

Table 2. Parameters employed and associated values.

Parameter	Value	Remarks
Optimizer	Stochastic Gradient Descent (SGD).	
Seed	0	reproducibility
Verbose	True	Enabling the printing of detailed information during training
Amp	True	Enabling automatic mixed precision for faster training
Fraction	1.0	Indicating the image fraction to process during training
Overlap_mask	True	Enabling the overlapping mask for mosaic augmentation
Mask_ratio	4	The ‘dropout’ is set to 0.0, indicating no dropout during training

learning rate	0.01	The final learning rate (lrf) is also 0.01
Momentum	0.937	The is set to
weight decay	0.0005	
Warmup Epochs	3.0	With a warmup momentum of 0.8 and a warmup bias learning rate of 0.1

One of the integral performance metrics deployed was the box loss metric. This metric assesses the proximity of the model’s predicted bounding boxes to the actual objects within the dataset. A value close to 1 for this metric indicates a consistent improvement in the model’s ability to generalise and accurately delineate the identified objects in the dataset.

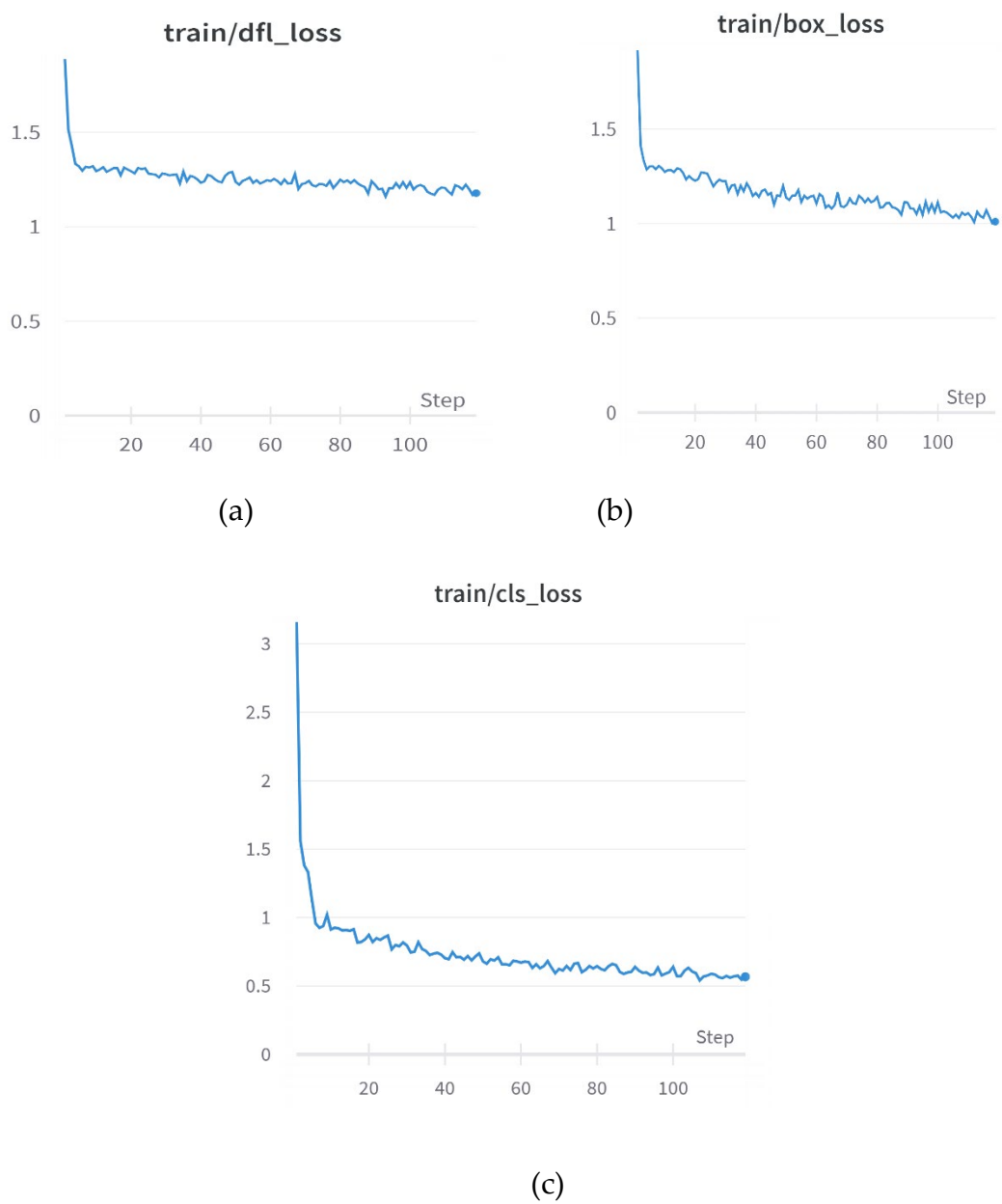


Figure 3. Outcomes of the training loss: (a) dfl loss, (b) box loss and (c) class loss.

The YOLO-v8 model underwent training with the drowsiness dataset for 600 epochs, with a batch size 16. The early stopping mechanism activated after 119 epochs due to a lack of observed improvement in the previous 50 epochs in adherence with the specified patient setting. The overall

training phase spanned roughly 0.3 hours on a Jupyter Notebook, with the best results manifesting at epoch 69.

4.1. Performance of the Yolov8

When subjected to cropped images of operatives, the accuracy of the YOLOv8 model in categorising the images into classes - Awake with PPE, awake without PPE, Drowsy with PPE, and Drowsy without PPE - was recorded as 64%, 43%, 83%, and 95% respectively. Interestingly, the model displayed greater proficiency in detecting classes characterised by drowsiness with or without PPE than classes depicting Awake with or without PPE. Notably, the class 'Awake without PPE' was misclassified as 'Drowsy without PPE' 57% of the time, indicating a propensity for false-positive detection in the model. This observation may be attributable to the imbalance in the dataset composition.

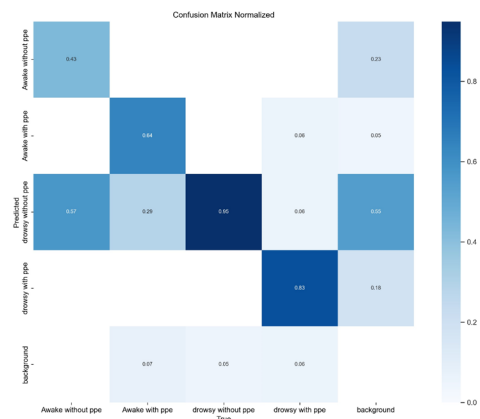
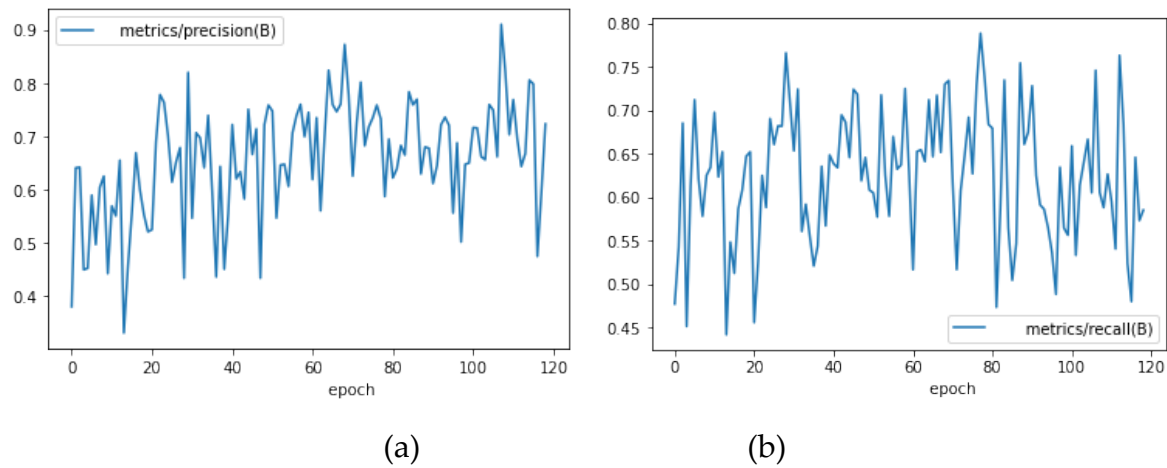
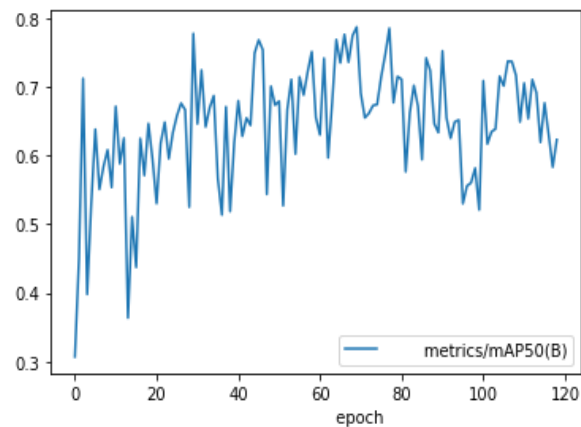


Figure 4. Confusion matrix for YOLOv8 accuracy for the dataset.





(c)

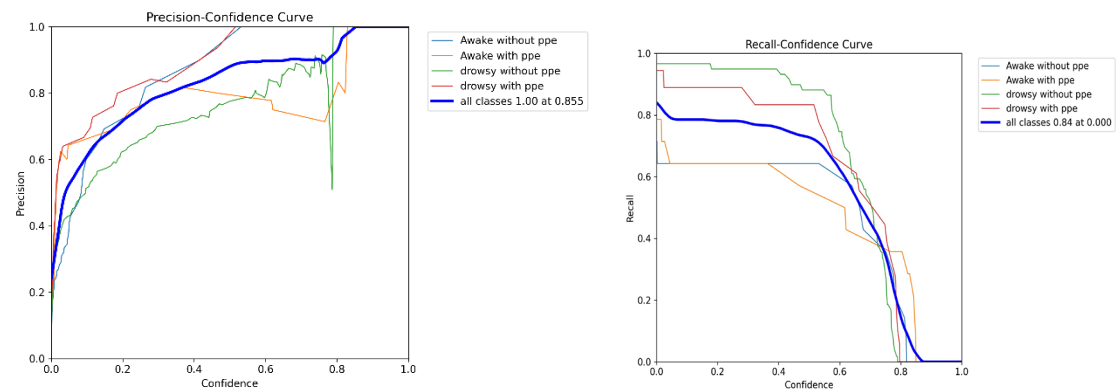
Figure 5. Performance of YOLO-v8 on the drowsiness dataset based on (a) precision, (b) recall and (c) mAP at the 50 IoU threshold.

The model’s performance, evaluated based on precision, recall, and mAP at the 50 IoU threshold, with the drowsiness dataset, is illustrated in Figure 2.

From the evaluation metrics obtained on the validation drowsiness dataset, as summarised in Table 1, we discern that the overall precision approximated 87%, with a recall of 73% and an mAP at the 50 IoU threshold of 78%. When delving into the class-wise performance, we note that the model detected ‘Awake without PPE’ with 96% precision, 64% recall, and 67% mAP at the 50 IoU threshold. The ‘Awake with PPE’ category had a precision of 80%, recall of 57%, and 68% mAP at the 50 IoU threshold. For the ‘Drowsy without PPE’ class, the precision was 77%, a recall was 88%, and the mAP at the 50 IoU threshold was 84%. The ‘Drowsy with PPE’ class exhibited a precision of 96%, recall of 83%, and 92% mAP at the 50 IoU threshold. The compelling contrast between the model’s proficiency in discerning drowsiness from being awake reiterates the significance of a balanced dataset in improving model performance.

Table 3. Validation outcomes on the drowsiness dataset.

Class	Metrics		
	Precision	Recall	mAP@50
All	0.87	0.73	0.78
Awake without PPE	0.96	0.64	0.67
Awake with PPE	0.80	0.57	0.68
Drowsy without PPE	0.77	0.88	0.84
Drowsy with PPE	0.96	0.83	0.92



(a) (b)

Figure 6. (a) YOLOv8 Precision-Confidence Curve (RCC), (b) YOLOv8 Recall-Confidence Curve (RCC).

The model’s performance correlation with respect to the precision of the dataset and corresponding confidence level is encapsulated in the Precision-Confidence Curve (PCC curve) depicted in Figure 5(a). This curve exhibits a gradual ascent until it peaks at 1.00 accuracy and 0.855 confidence, maintaining this level after that. The Recall-Confidence Curve (RCC) in Figure 5(b) elucidates the inverse relationship between the recall of the dataset and confidence. The distinct behaviours observed in both these curves further underscore the complexity and intricacies involved in achieving high precision and recall in object detection tasks. Both metrics have been shown to have a relationship that is inversely proportional to one another.



Figure 7. Drowsiness datasets Actual labels.

Figure 8 visually juxtaposes actual labels from the validation dataset with the predicted labels post-training, further illuminating the efficacy of our model. According to the outcomes of the tests and validations, drowsiness detection was carried out well. However, the awake with PPE had a noticeable result, which is to be expected given the small sample size of the dataset.



Figure 8. Drowsiness datasets predicted labels.

4.2. Discussion

This study’s objective was to discern the viability of the YOLOv8 model in detecting drowsiness among construction operatives with significant accuracy and in real time. Our findings reveal promising outcomes, with the YOLOv8 model displaying remarkable precision in the images and an inference speed reaching 0.4ms for preprocessing, 7.5ms for inference, 0.0ms for loss calculation, and 1.7ms for postprocessing per image in the drowsiness dataset. Further, the model achieved a notable mean average precision (mAP) of 92% in the drowsiness detection task.

Despite these encouraging results, examining the model’s performance exposes a shortcoming in handling imbalanced classes. The underrepresented class, ‘Awake with PPE’, was detected with high precision, but comparatively, it demonstrated a lower recall and mAP in the testing dataset. This disparity in performance across various classes reinforces the need for a balanced dataset and well-chosen hyperparameters to ensure optimal performance in deep learning-based object detection tasks [33].

The performance of the YOLOv8 model was also compared with similar works in the field, as indicated in Table 2. Our mAP of YOLOv8 detecting drowsiness was 78%, which compares favourably with other studies employing different methodologies for similar or related object detection tasks. For instance, Wang et al. [43] employed the R-CNN model to detect safety hazards and reported an mAP of 92.55%. On the other hand, Lee et al. [34] utilised the YOLOACT model for detecting construction worker presence, hardhat usage, and safety vest usage, with reported mAPs of 64.3%, 77.2%, and 62.3%, respectively. Our results substantiate the potential of YOLOv8 for real-time detection tasks, with the added advantage of computational efficiency.

Table 4. A general comparison between the YOLOv8 model and other related work.

Reference	Domain	Methodology	mAP (%)
Proposed	Drowsiness	Yolo-v8	78
Lee et al. (2023)	Worker	YOLOACT	64.3
Lee et al. (2023)	Hardhat	YOLOACT	77.2
Lee et al. (2023)	Safety vest	YOLOACT	62.3

The comprehensive nature of our approach, spanning from data collection to model training, validation, and comparative analysis, underpins the reliability of our findings. It is noteworthy that YOLOv8’s technical superiority is emphasised in the study, providing a detailed insight into its performance across different classes.

Implications of the Study

Drowsiness as an Indicator of Fatigue

Given the propensity of most fatigue measurements to be subjective, an algorithm to identify drowsiness helps to easily measure workers’ readiness and cognition to be involved in construction tasks. This also helps workers to be aware of the effects of sleep deprivation and what they can do on a personal level to sleep better and be more physically ready for laborious construction tasks.

Enhanced Construction Health, Safety and Well-Being

With an improved method of identifying drowsiness, risks associated with fatigue are reduced as against using subjective measures which workers can easily manipulate. This helps avoid risk to the worker, the client and the contractor.

Although the created computer vision approach cannot 100% identify drowsy workers, it offers project managers various advantages in their daily work. To begin with, safety behaviour can be observed without disrupting those who are at work. Secondly, a variety of operating areas can be

simultaneously observed, which can cut down on the expense and duration of inspections. It provides a way to lessen hazards on site as a system in place to monitor persons with physical fatigue from being a danger to themselves and their teammates.

5. Conclusions and Future Directions

The intricacies and dangers of construction sites necessitate diligent safety management procedures. The prevailing practice of manual surveillance and analysis is labour-intensive and prone to errors. Even more worrisome is the fact that warnings of potential safety dangers may be subjective and occur at an inconvenient time, which might ultimately result in tragic effects. Consequently, computer vision and deep learning algorithms have emerged as transformative tools, driving improved competency and efficiency in safety management across construction sites. The findings of this research are particularly significant in two respects. Firstly, this pioneering work may pave the way towards safer and smarter construction sites by enabling real-time identification and intervention for drowsy operatives. Second, because this study was one of the initial deployments of the most recent YOLOv8 model in the construction sector, it provides vital insights to the field of computer vision research regarding the resilience and usability of this advanced approach.

However, this study was not without its limitations. A critical compromise was made in image resizing to standardise the location of bounding boxes during dataset annotation. As a result, the original dataset was scaled down to a uniform size of 416x416 pixels, inadvertently leading to some low-quality images. While necessary for model training, this modification poses challenges for real-world implementation where high-resolution images may be crucial for accurate detections.

The enhancements to the proposed system will primarily revolve around improving drowsiness detection. This can be achieved by augmenting the training datasets with more diverse data, thus enabling the model to learn from a richer variety of instances. Further research could also consider implementing the model on a video-based inspection system and developing a timed alarm system to proactively alert supervisors of potential safety risks. Given the supervised learning techniques employed in this study, exploring the possibility of combining supervised and unsupervised learning methods could result in the development of a more intelligent system, heralding the beginning of a new era of safety management in the construction sector. Ultimately, the study has provided a springboard for applying advanced deep learning models, specifically YOLOv8, in construction safety management. It underscores the transformative potential of AI in ensuring occupational safety, and the findings contribute both to the practical field of construction safety and the academic realm of computer vision research.

Author Contributions: “Conceptualization, A.O. and D.O.; methodology, D.O.; software, D.O.; validation, A.O., D.O. and A.S.; formal analysis, D.O.; investigation, A.O.; resources, A.O.; data curation, S.A.; writing—original draft preparation, A.O.; writing—review and editing, A.O.; visualization, D.O.; supervision, I.M.; project administration, A.O.; funding acquisition, A.O. All authors have read and agreed to the published version of the manuscript.”

Funding: This research was funded by the National Research Foundation (NRF) under Grant Number 129953.

Data Availability Statement: Data available upon request.

Acknowledgments: The funding for this research has been provided by the National Research Foundation (NRF) under Grant Number 129953. The opinions and conclusions expressed in this work belong to the authors and may not necessarily be attributed to the NRF. The research is funded and conducted as a collaborative effort at the Centre of Applied Research and Innovation in the Built Environment (CARINBE), which is affiliated with the Faculty of Engineering and the Built Environment at the University of Johannesburg.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Luo, H.; Wang, M.; Wong, P. K. Y.; Tang, J.; Cheng, J. C. P. Vision-Based Pose Forecasting of Construction Equipment for Monitoring Construction Site Safety. *Lect. Notes Civ. Eng.* **2021**, 98 (September 2020), 1127–1138. https://doi.org/10.1007/978-3-030-51295-8_78.
2. Onososen, A. O.; Musonda, I.; Onatayo, D.; Tjebane, M. M.; Saka, A. B.; Fagbenro, R. K. Impediments to Construction Site Digitalisation Using Unmanned Aerial Vehicles (UAVs). *Drones* **2023**, 7 (1), 45. <https://doi.org/10.3390/drones7010045>.
3. Powell, R.; Copping, A. Sleep Deprivation and Its Consequences in Construction Workers. *J. Constr. Eng. Manag.* **2010**, 136 (10), 1086–1092. [https://doi.org/10.1061/\(asce\)co.1943-7862.0000211](https://doi.org/10.1061/(asce)co.1943-7862.0000211).
4. Ansaripour, A.; Heydariaan, M.; Kim, K.; Gnawali, O.; Oyediran, H. Applied Sciences ViPER + : Vehicle Pose Estimation Using Ultra-Wideband Radios for Automated Construction Safety Monitoring. *Appl. Sci.* **2023**, 13 (1581).
5. NCS. Fatigue in Safety Critical Industries — Impacts, Risks & Recommendations <https://www.nsc.org/in-the-newsroom/69-percent-of-employees-many-in-safety-critical-jobs-are-tired-at-work-says-nsc-report> (accessed 2023 -06 -25).
6. Sedighi Maman, Z.; Alamdar Yazdi, M. A.; Cavuoto, L. A.; Megahed, F. M. A Data-Driven Approach to Modeling Physical Fatigue in the Workplace Using Wearable Sensors. *Appl. Ergon.* **2017**, 65, 515–529. <https://doi.org/10.1016/j.apergo.2017.02.001>.
7. Zhang, K.; Li, X. Human-Robot Team Coordination That Considers Human Team Fatigue. *Int. J. Adv. Robot. Syst.* **2014**, 11 (9), 1–9. <https://doi.org/10.5772/58228>.
8. Yu, Y.; Guo, H.; Ding, Q.; Li, H.; Skitmore, M. An Experimental Study of Real-Time Identification of Construction Workers' Unsafe Behaviors. *Autom. Constr.* **2017**, 82 (July), 193–206. <https://doi.org/10.1016/j.autcon.2017.05.002>.
9. Ray, S. J.; Teizer, J. Real-Time Construction Worker Posture Analysis for Ergonomics Training. *Adv. Eng. Informatics* **2012**, 26 (2), 439–455. <https://doi.org/10.1016/j.aei.2012.02.011>.
10. Wójcik, B.; Źarski, M.; Książek, K.; Miszczak, J. A.; Skibniewski, M. J. Hard Hat Wearing Detection Based on Head Keypoint Localization. **2021**.
11. Mariam, A. T.; Olalusi, O. B.; Haupt, T. C. A Scientometric Review and Meta- Analysis of the Health and Safety of Women in Construction : Structure and Research Trends. *J. Eng. Des. Technol.* **2021**, 19 (2), 446–466. <https://doi.org/10.1108/JEDT-07-2020-0291>.
12. Fang, Q.; Li, H.; Luo, X.; Ding, L.; Luo, H.; Li, C. Computer Vision Aided Inspection on Falling Prevention Measures for Steeplejacks in an Aerial Environment. *Autom. Constr.* **2018**, 93 (May), 148–164. <https://doi.org/10.1016/j.autcon.2018.05.022>.
13. Gheisari, M.; Rashidi, A.; Esmaili, B. Using Unmanned Aerial Systems for Automated Fall Hazard Monitoring. *Constr. Res. Congr. 2018 Saf. Disaster Manag. - Sel. Pap. from Constr. Res. Congr. 2018* **2018**, 2018-April (June), 62–72. <https://doi.org/10.1061/9780784481288.007>.
14. Kim, D.; Lee, S.; Kamat, V. R. Proximity Prediction of Mobile Objects to Prevent Contact-Driven Accidents in Co-Robotic Construction. *J. Comput. Civ. Eng.* **2020**, 34 (4), 04020022. [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000899](https://doi.org/10.1061/(asce)cp.1943-5487.0000899).
15. Srinivas Aditya, U. S. P.; Singh, R.; Singh, P. K.; Kalla, A. A Survey on Blockchain in Robotics: Issues, Opportunities, Challenges and Future Directions. *J. Netw. Comput. Appl.* **2021**, 196 (March 2021), 103245. <https://doi.org/10.1016/j.jnca.2021.103245>.
16. Tjebane, M. M.; Musonda, I.; Okoro, C. Organisational Factors of Artificial Intelligence Adoption in the South African Construction Industry. *Front. Built Environ.* **2022**, 8 (March). <https://doi.org/10.3389/fbuil.2022.823998>.
17. Bowen, P.; Peihua Zhang, R.; Edwards, P. An Investigation of Work-Related Strain Effects and Coping Mechanisms among South African Construction Professionals. *Constr. Manag. Econ.* **2021**, 39 (4), 298–322. <https://doi.org/10.1080/01446193.2020.1870045>.
18. Seo, J.; Han, S.; Lee, S.; Kim, H. Computer Vision Techniques for Construction Safety and Health Monitoring. *Adv. Eng. Informatics* **2015**, 29 (2), 239–251. <https://doi.org/10.1016/j.aei.2015.02.001>.
19. Luo, H.; Wang, M.; Wong, P. K. Y.; Tang, J.; Cheng, J. C. P. Construction Machine Pose Prediction Considering Historical Motions and Activity Attributes Using Gated Recurrent Unit (GRU). *Autom. Constr.* **2021**, 121 (October 2020), 103444. <https://doi.org/10.1016/j.autcon.2020.103444>.
20. Souto Maior, C. B.; Santana, J. M.; Nascimento, L. M.; Macedo, J. B.; Moura, M. C.; Isis, D. L.; Droguett, E. L. Personal Protective Equipment Detection in Industrial Facilities Using Camera Video Streaming. *Saf. Reliab. - Safe Soc. a Chang. World - Proc. 28th Int. Eur. Saf. Reliab. Conf. ESREL 2018* **2018**, No. October 2019, 2863–2868. <https://doi.org/10.1201/9781351174664-359>.
21. Liang, C. J.; Kamat, V. R.; Menassa, C. C. Teaching Robots to Perform Construction Tasks via Learning from Demonstration. In *Proceedings of the 36th International Symposium on Automation and Robotics in Construction, ISARC 2019; 2019*; pp 1305–1311. <https://doi.org/10.22260/isarc2019/0175>.

22. Akinosho, T. D.; Oyedele, L. O.; Bilal, M.; Ajayi, A. O.; Delgado, M. D.; Akinade, O. O.; Ahmed, A. A. Deep Learning in the Construction Industry: A Review of Present Status and Future Innovations. *J. Build. Eng.* **2020**, *32*, 101827. <https://doi.org/10.1016/j.jobe.2020.101827>.
23. Ibrahim, A.; Nnaji, C.; Namian, M.; Koh, A.; Techera, U. Investigating the Impact of Physical Fatigue on Construction Workers' Situational Awareness. *Saf. Sci.* **2023**, *163* (January), 106103. <https://doi.org/10.1016/j.ssci.2023.106103>.
24. Malomane, R.; Musonda, I. The Opportunities and Challenges Associated with the Implementation of Fourth Industrial Revolution Technologies to Manage Health and Safety. **2022**.
25. FinancialTimes. Qatari World Cup Chief Says 400 Workers Killed during Construction. *ft.com*. 2022.
26. Onososen, A. O.; Musonda, I.; Ramabodu, M.; Tjebane, M. M. Resilience in Construction Robotics and Human-Robot Teams for Industrialised Construction. In *The Twelfth International Conference on Construction in the 21st Century (CITC-12 Amman, Jordan | May 16-19, 2022; 2022*; pp 537–546.
27. Nath, N. D.; Behzadan, A. H.; Paal, S. G. Deep Learning for Site Safety: Real-Time Detection of Personal Protective Equipment. *Autom. Constr.* **2020**, *112* (January), 103085. <https://doi.org/10.1016/j.autcon.2020.103085>.
28. Anwer, S.; Li, H.; Antwi-Afari, M. F.; Umer, W.; Mehmood, I.; Al-Hussein, M.; Wong, A. Y. Test-Retest Reliability, Validity, and Responsiveness of a Textile-Based Wearable Sensor for Real-Time Assessment of Physical Fatigue in Construction Barbenders. *J. Build. Engineering* **2021**, *44* (1). <https://doi.org/https://doi.org/10.1016/J.JOBE.2021.103348>.
29. Namian, M.; Albert, A.; Feng, J. Effect of Distraction on Hazard Recognition and Safety Risk Perception. *J. Constr. Eng. Manag.* **2018**, *144* (4). <https://doi.org/04018008>. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001459](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001459).
30. Soltani, M. M.; Zhu, Z.; Hammad, A. Skeleton Estimation of Excavator by Detecting Its Parts. *Autom. Constr.* **2017**, *82*. <https://doi.org/https://doi.org/10.1016/j.autcon.2017.06.023>.
31. Cha, Y.-J.; Choi, W.; Büyüköztürk, O. Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. *Comput. Civ. Infrastruct. Eng.* **2017**, *32*, 361–378. <https://doi.org/https://doi.org/10.1111/mice.12263>.
32. Yokoyama, S.; Matsumoto, T. Development of an Automatic Detector of Cracks in Concrete Using Machine Learning. *Procedia Eng.* **2017**, *171*, 1250–1255. <https://doi.org/https://doi.org/10.1016/J.PROENG.2017.01.418>.
33. Nepal, U.; Eslamiat, H. Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs. *Sensors* **2022**, *22* (2). <https://doi.org/10.3390/s22020464>.
34. Lee, Y.-R.; Jung, S.-H.; Kang, K.-S.; Ryu, H.-C.; Ryu, H.-G. Deep Learning-Based Framework for Monitoring Wearing Personal Protective Equipment on Construction Sites. *J. Comput. Des. Eng* **2023**, *10*, 905–917. <https://doi.org/https://doi.org/10.1093/jcde/qwad019>.
35. Onatayo, D. A.; Srinivasan, R. S.; Shah, B. Ultraviolet Radiation Transmission in Building's Fenestration: Part II, Exploring Digital Imaging, UV Photography, Image Processing, and Computer Vision Techniques. **2023**.
36. Karlsson, J.; Strand, F.; Bigun, J.; Alonso-Fernandez, F.; Hernandez-Diaz, K.; Nilsson, F. Visual Detection of Personal Protective Equipment and Safety Gear on Industry Workers. **2023**, 395–402. <https://doi.org/10.5220/0011693500003411>.
37. Aich, S.; Joo, M. il; Kim, H. C.; Park, J. Improvisation of Classification Performance Based on Feature Optimization for Differentiation of Parkinson's Disease from Other Neurological Diseases Using Gait Characteristics. *Int. J. Electr. Comput. Eng.* **2019**, *9* (6), 5176–5184. <https://doi.org/10.11591/ijece.v9i6.pp5176-5184>.
38. Karystianis, G.; Adily, A.; Schofield, P.; Knight, L.; Galdon, C.; Greenberg, D.; Jorm, L.; Nenadic, G.; Butler, T. Automatic Identification of Mental Health Disorders from Domestic Violence Police Narratives. *J. Med. Internet Res* **2018**, *20*. <https://doi.org/https://doi.org/10.2196/11548>.
39. Aich, S.; Joo, M.; Kim, H.-C.; Park, J. Improvisation of Classification Performance Based on Feature Optimization for Differentiation of Parkinson's Disease from Other Neurological Diseases Using Gait Characteristics. *Int. J. Electr. Comput. Eng* **2019**, *9*, 5176–5184. <https://doi.org/https://doi.org/10.11591/ijece.v9i6.pp5176-5184>.
40. Cromsjö, M.; Hallonqvist, L. Detection of Safety Equipment in The Manufacturing Industry Using Image Recognition. **2021**.
41. Onatayo, D. A.; Srinivasan, R. S.; Shah, B. Ultraviolet Radiation Transmission in Buildings' Fenestration: Part I, Detection Methods and Approaches Using Spectrophotometer and Radiometer. *Buildings* **2023**, *13* (7), 1670. <https://doi.org/10.3390/buildings13071670>.
42. Aboah, A.; Wang, B.; Bagci, U.; Adu-Gyamfi, Y. Real-Time Multi-Class Helmet Violation Detection Using Few-Shot Data Sampling Technique and YOLOv8. **2023**, 5349–5357.
43. Wang, M.; Wong, P. K.-Y.; Luo, H.; Kumar, S.; Delhi, V.-S. ; Cheng, J. C.-P. Predicting Safety Hazards Among Construction Workers and Equipment Using Computer Vision and Deep Learning Techniques. In

Presented at the 36th International Symposium on Automation and Robotics in Construction, Banff, AB, Canada; 2019. <https://doi.org/https://doi.org/10.22260/ISARC2019/0054>.

44. Aboah, A.; Wang, B.; Bagci, U.; Adu-Gyamfi, Y. Real-Time Multi-Class Helmet Violation Detection Using Few-Shot Data Sampling Technique and YOLOv8. **2023**.
45. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y. M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. **2022**, No. July. <https://doi.org/10.48550/arXiv.2207.02696>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.