Computer Vision Enabled Drowning Detection System

Upulie Handalage
Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it18107074@my.sliit.lk

Nisansali Nikapotha
Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it18111866@my.sliit.lk

Chanaka Subasinghe
Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it18110012@my.sliit.lk

Tereen Prasanga
Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it18087734@my.sliit.lk

Thusithanjana Thilakarthna
Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
thusithanjana.t@sliit.lk

Dharshana Kasthurirathna
Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
dharshana.k@sliit.lk

2021 3rd International Conference on Advancements in Computing (ICAC) | 978-1-6654-0862-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICAC54203.2021.9671126

***Abstract* -Safety is paramount in all swimming pools. The current systems expected to address the problem of ensuring safety at swimming pools have significant problems due to their technical aspects, such as underwater cameras and methodological aspects such as the need for human intervention in the rescue mission. The use of an automated visual-based monitoring system can help to reduce drownings and assure pool safety effectively. This study introduces a revolutionary technology that identifies drowning victims in a minimum amount of time and dispatches an automated drone to save them. Using convolutional neural network (CNN) models, it can detect a drowning person in three stages. Whenever such a situation like this is detected, the inflatable tube-mounted self-driven drone will go on a rescue mission, sounding an alarm to inform the nearby lifeguards. The system also keeps an eye out for potentially dangerous actions that could result in drowning. This system's ability to save a drowning victim in under a minute has been demonstrated in prototype experiments' performance evaluations.**

Keywords—Drowning, Lifeguard system, Object detection, Computer vision, Pose estimation, Drone, Convolutional Neural Network (CNN)

I. Introduction

Drowning is the third most significant cause of accidental injury globally, according to the World Health Organization (WHO) [1]. Responsible for 7% of all injury-related fatalities, an estimated 320 000 people drown each year [2]. An average of 855 persons perished each year in Sri Lanka due to drowning per year, resulting in a drowning rate of 4.4 fatalities per 100,000 people [1]. Drowning may happen in various settings, including bathtubs, natural water bodies, and swimming pools. According to National Vital Statistics System (NVSS), swimming pools account for approximately 16 percent of all drowning deaths, implying a dangerous link between pool swimming and mortality [2], [3].

According to studies, lifeguards may not be adequately trained to deal with a drowning incident [2]. Whether it's due to a lack of training or a failure to spot a drowning victim quickly enough, the result of a life and death scenario may change instantaneously.

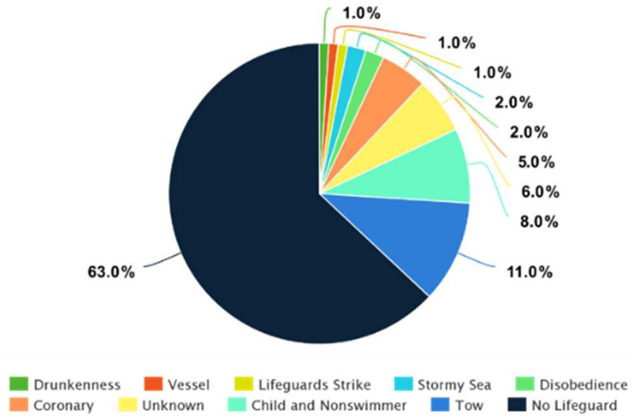

Fig. 1. Relative percentage of drowning causes according to lifeguards' reports

Apart from drowning, those who disobey pool laws and regulations cause discomfort to others while causing serious health problems for themselves with dangerous actions such as drinking and running around the pool [4], [5]. Fig.1 shows how lifeguards claim that lack of care is the most probable cause of drowning, among the other reasons that intoxication appears to be present [6], [7].

Considering the possibility that attempts to save lives in the water using traditional methods fail, it is clear that an intelligent system is needed. The system described in this paper uses computer vision to detect and rescue drowning victims to ensure the safety of pools. The system uses computer vision along with automated electronic equipment to immediately rescue and protect the lives of swimmers in the pool. By integrating the camera above the water surface, it can recognize struggling motions before a fatality occurs. The camera's location captures a complete view of the facility, including swimmers, wanderers, and occupied objects. Swimmers are individually identified using object detection, noise cancellation and individually tracked using deep learning technologies to identify a possible drowning. On detection, the location coordinates of the drowning person are immediately calculated based on the ground coordinates (a grid system linked to x and y blocks) and sent to an autonomous custom drone while sounding an alarm,

978-1-6654-0862-2/21/$31.00 ©2021 IEEE

informing that more security measures should be taken. In addition, the detection of hazardous activities through computer vision concepts and posture detection in the facility's swimming pool ensures the safety and well-being of swimmers. This goal is achieved by using Firebase Cloud Messaging (FCM), the primary notification source to authorized personnel if a dangerous activity is detected.

The structure of the article is as follows. Firstly, the literature survey discusses the currently available existing systems and technologies that have similar targets using both software and hardware-based technologies. Secondly, the methodology section explains the steps of how the system tried to solve this problem. Next, the results and discussion section analyze the main experimental results. The final section discusses the advantages and disadvantages of and possible work to improve the system in the future.

II. Literature Survey

Vision-based systems and wearable sensor-based systems are two types of existing drowning detection technologies. Vision-based technologies are further subdivided into those that use underwater cameras [8], [9], [10], [11], [12], [13], [14] and those that use above-water cameras. Underwater cameras have the drawback of missing the early struggle above the water. Early on, failure to recognize a drowning scene could result in a longer rescue time, which is a significant issue to consider in a time-critical emergency. The main disadvantage of a wearable-based system is the discomfort of use, which may lead to younger children seeking to alleviate the discomfort by removing the device, which is an unsubstantiated theory [15].

A. *Object Detection Using Different Techniques*

It is claimed that the usage of Convolutional Neural Network (CNN) architecture in Deep Neural Networks (DNNs) has added a significant shift in learning more complicated, informative characteristics in images as compared to the older techniques [16]. Furthermore, further optimized models such as Fast R-CNN, Faster R-CNN, and YOLO have been constructed since the region-based convolutional neural network (R-CNN) architecture proposal. Fast R-CNN, which improves bounding box (BB) regression and classification [17]; Faster R-CNN [18], which generates area suggestions using an extra sub-network [18]; and YOLO, which detects objects using a fixed-grid regression [19], are all faster than R-CNN. Bounding box regression is used to recognize generic objects based on basic CNN architectures. Local contrast enhancement and pixel-level segmentation, on the other hand, are used to recognize salient objects [20]. The techniques used in detecting objects under this chapter will be crucial as they establish the groundwork for the methodologies used to identify drowning and hazardous activities.

B. *Drowning Detection And Tracking*

To avoid drowning events utilizing an alert system, Alshbatat et al. [10] proposed an integrated vision-based monitoring system consisting of a Raspberry Pi, two Pixy cameras, and an Arduino Nano board. They employed two cameras to detect and monitor swimmers by measuring their positions, and the swimmers were obliged to wear passive yellow vests. NEPTUNE [21], is another unique technology that uses statistical image processing [22] of video sequences to detect drowning victims as soon as possible. The equations utilized in detecting near-drowning victims are based on the variables created by statistical image processing. Another system called VIBE [15] uses background extraction to detect and track drowning victims and updates the motion area by taking the frame difference using the VIBE algorithm, which primarily evaluates the swimmers' positions when making judgments. How-Lung et al. [23] examine some difficulties in spotting drowning victims in a watery environment and offer an automatic detection surveillance system. The key obstacles in the aquatic environment, according to the authors, are water ripples and splashes, as well as background movements of the reflective zones. When it comes to recognizing swimmers, occlusions are also mentioned as a challenging difficulty. Their proposed solution is an algorithm that takes into account all of these issues and detects water crises in complex aquatic environments [24], [25].

C. *Activity Detection Using Computer Vision*

Current work on human motion prediction has been focused on two independent but complementary sub-tasks, according to Anand Gopalkrishnan [26]. 1) Short-term motion prediction, which is quantitatively evaluated by measuring the mean squared error (MSE) over a short period, and 2) long-term motion prediction, qualitatively evaluated by visual inspections of samples over a long period. Short-term models would be valuable in motion tracking applications because these jobs are applicable in several domains of work. On the other hand, long-term models might be valuable for creating computer graphic tools due to their broad applicability. Additionally, both models could be useful in human gait analysis, kinematics research, and human-computer interaction.

III. Methodology

The system explained in this paper includes three main functions: detecting drowning victims, sending drones to victims, and detecting dangerous activities. The drowning detection component detects drowning victims through a custom CNN model, which detects drowning in three stages and immediately informs the user through an audio alert. The second component is the rescue drone, activated according to the drowning detection command and sent to the victim's location coordinates. This procedure uses a custom-configured x and y coordinate block system to link to ground GPS coordinates. At the same time, potentially dangerous activities, including running around the swimming pool and drinking, will be notified to authorized personnel in the premises through mobile alarms by the hazard detection component. This will prompt authorized personnel (including lifeguards) to make responsible decisions. Fig.2 shows the main procedure of the system.

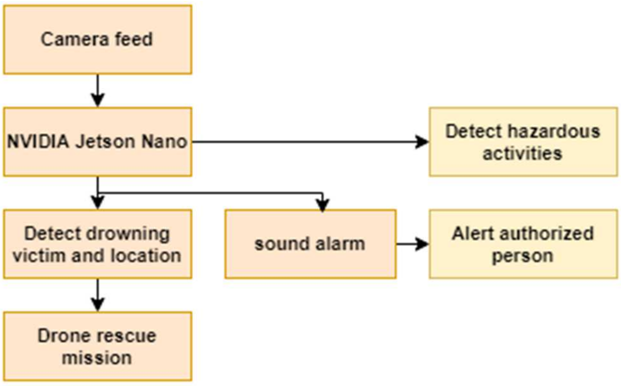

Fig.1. High overview of the system

A. Drowning Detection And Tracking

1. *Creation of the data set:* Due to the lack of an existing aquatic human body parts data set, a data set containing 5000 images were constructed. All images in the dataset contain at least one or more swimmers in the water
   - *Image Collection*: The primary source of data collection is the induction of actors and the collection of videos in real-time. The secondary source of data collection is the Internet, using specific keywords, such as "swimmer", "swimming", "drowning", "drowning in a swimming pool",
   - *Image Labelling:* LabelImg – a graphical tool implemented in Python, is used to mark the image. Each image is labelled by creating arbitrary bounding boxes and predefined labels in YOLO format. The predefined tags used in image annotation are "Not_drowning", "Drowning_stage_1", "Drowning_stage_2", and "Drowning_stage_3".
2. *Model Creation:* Use Google Colab to create and train models and get weight files every 100 iterations. The created model is then implemented on the NVIDIA Jetson Nano board, which runs on the Quad-core ARM CortexA57 processor. The main reason for using NVIDIA Jetson is to run multiple neural models parallel without complications and with a limited budget.

First, swimmers in the pool are detected using an overhead camera and are kept track using the DeepSORT algorithm. YOLO is used to detect objects by locating one or more objects in the image and sorting each object. Yolo works well with a good resolution of entry compared to other models [20]. Most of the problems in the detection and monitoring of swimming players are occlusal, scale changes, changes of appearance. These problems can be overcome using YOLO [27]. The location of the tracked swimmer is also obtained using a predefined coordinate system.

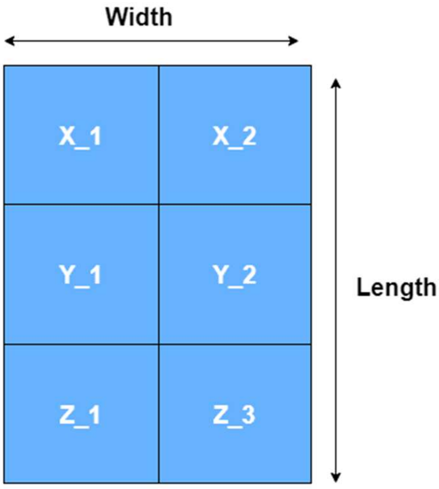

Fig.2. Sample predefined coordinate system

Fig.3 shows the custom predefined coordinate system for the swimming pool. Initially, the detection of swimmers is tracked while using a predefined coordinate system to obtain their position coordinates. At the same time, it will also check whether the swimmer has entered any drowning stage. If such an event is identified, the camera will track video clips in real-time to detect drowning victims. Using CNN, the human detection algorithm will use the image frames to identify the drowning person. Once the location block of the drowning victim is correctly identified, an audible alarm will sound to notify authorized personnel of the event.

The images in the data set must include at least one drowning person to identify the chart as a drowning situation. A person's posture and movement can quickly identify a drowning victim. Although it is easy to recognize, one of the most common exercises is the "vertical ladder" exercise, which imitates the movement of a person climbing a ladder in a vertical movement [28].

TABLE I. MAIN CLASSES USED IN IDENTIFYING A DROWNING VICTIM

Class	Features
Drowning_stage_1	Head is above the water
Drowning_stage_2	Half of the head is underwater, and hand gestures are in the climbing ladder motion
Drowning_stage_3	Head is underwater, and hand gestures are in the climbing ladder motion
Not_drowning	Regular swimming and floating motions

Table I describes the four categories used to classify swimmers: Drowning_stage_1, Drowning_stage_2, Drowning_stage_3, and Not_drowning. In the event of drowning, a frame pattern of stage 1 to stage 2 and then to stage 3 can be seen. Fig.4 consists of three images, each 743 pixels and 243 pixels in size. According to the climbing ladder motion, people who switched between stages 2 and 3 were identified as drowning victims. Finally, the location block of the drowning victim is passed to the drone.

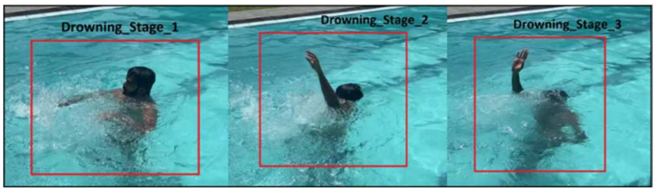

Fig.3. Three primary drowning stages

B. Identification Of Hazardous Activities

In addition, the system can also analyze all visible activities in the pool to ensure the swimmer's safety. Continuous monitoring ensures that dangerous activities are not carried out on the premises. This is accomplished by notifying authorized personnel via mobile alarms when a dangerous activity occurs.

The process of identifying hazardous activities at a location is initiated by collecting hazardous and non-hazardous activity data sets. The data set is collected in a pool that contains people hanging out in an environment. The identification process after the alarm notification can be divided into four steps.

1. *Masking and noise extraction*: Due to the camera arrangement, the noise obtained from the water surface at the site is concentrated in the water waves during the day. As shown in Fig.5, which is of size 1280 x 720 pixels, a fundamental step of masking the image is performed to remove the image's noisy areas. Due to the non-stationary nature of the camera, all frames have static X and y coordinates as masking points.

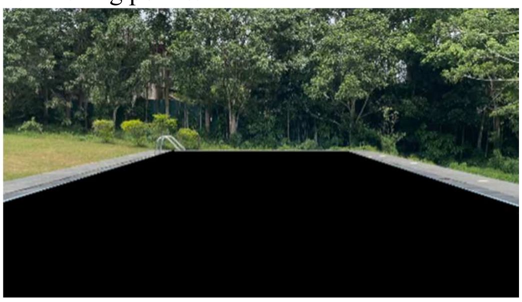

Fig.4. Masked image for hazardous activity prediction

2. *Skeleton sketching*: Skeleton sketching is done using OpenPose - a real-time multi-person keypoint detection library [29], [30] for pose estimation, which is then used to draw skeleton-based human figures to recognize a person's pose in real-time [31]. The poses recognized by pose estimation uses a combination of DNN models to successfully approximate a complex nonlinear mapping function from a random image of a person to match the position, as shown in Fig.6. Each of the images is of size 1280 x 720 pixels due to the static nature of the camera placement.

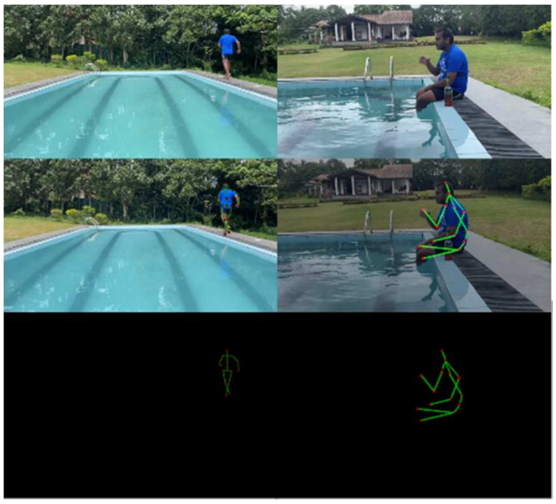

Fig.5. Steps in retrieving the CNN input

3. *Labelling images using CNN:* Using transfer learning, a YOLOv3 model [19] was improved to create a custom model. The sketched dark backgrounded images identified through the skeleton sketching process are integrated into the customized CNN model, identifying the categories (hazardous and non-hazardous activities). The training is carried out in two batches at the Learning Rate of 0.001, and the data set maintains a set of 1000 images for each class labelled for training.
4. *Notifying the authorized personnel using mobile alerts:* A status alert is sent to one of the authorized personnel devices if a frame is classified as hazardous. This, in turn, will notify the regarded personnel to initiate necessary precautions and measures. The notification signal would be in the form of a Firebase push notification, as it is swift and easily comprehensive. The Android service [32] is compatible with any Android mobile running version from 4.4 (KitKat) and above.

IV. RESULTS AND DISCUSSION

A. Drowning Detection and Tracking Results

The YOLO detection algorithm [19] uses 416 X 416 as its input dimensions. The drowning victims are detected in three stages using a YOLO-based detection technique. Even though the swimmer stayed underwater for an extended period, the DeepSORT algorithm [33] could keep track of them. The mentioned Fig.7 depicts the model's performance (False Positives and True Positives only) with 500 images.

Fig.6. Detection results for drowning

Table II shows the test accuracies of the model as a percentage of classified photos, which was produced using a model with 500 testing images. Equation (1) is used to calculate the model's accuracy.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

TP - True Positives
FP - False Positives
TN - True Negatives
FN - False Negatives

TABLE II. ACCURACIES OF THE TESTING DATASET

Accuracy Variables	Count
TP	220
TN	208
FP	42
FN	30
Total Accuracy	85.6%

B. Hazardous Activities

Because of the noise elimination via picture masking, the posture estimate accuracy was greatly improved. To allow the pose estimation algorithm to make more radical judgments, the default threshold value for the OpenPose body parts heat map was changed from 0.2 to 0.1. Although frame-by-frame identifications were only identified with a probability of 53% due to the threshold adjustment, the total system, which examined a frame in real-time, was able to identify a hazardous activity with much greater ease within 60 seconds, with a mean accuracy of 91.4 percent, after the threshold was changed. Fig.8 illustrates a few of the postures recognized of various positions.

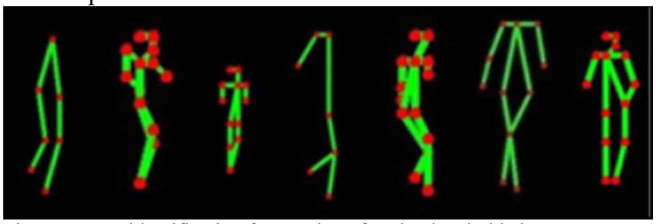
Fig.7. Posture identification for running after the threshold change

A close examination of the misclassified postures among the testing sets revealed that a posture was more prone to misclassification as it approached the far end of the camera, indicating the need for a secondary camera to improve accuracy and confirm the true positives from the primary camera, as shown in Table II. Although employing a higher-quality camera to fix this problem is a good idea, the requisite hardware and the near-real-time CNN techniques used to detect further objects may not be up to standard at present.

CONCLUSION AND FUTURE WORK

This computer vision-enabled automated drone-based lifeguard system consists of three main components, i.e., the drowning detection, the rescuing drone, and the hazardous activity detection. All three components combined will create a system capable of detecting drowning victims, dispatching an inflatable tube using a drone (as depicted in Fig.9) and detecting hazardous activities—eventually becoming an entity that could assist a lifeguard. The system is accessible to its primary user, presumably a pool owner or a lifeguard, in the form of an interface with a sound alarm and an android mobile service that holds the capabilities of receiving Firebase notifications.

Confined with a few of the hardware limitations, such as the use of a single camera and the Jetson Nano at the presence of better-quality hardware, could affect the speed and accuracy of the overall system is becoming a state-of-the-art. This limitation could be omitted with the use of multiple cameras that could be placed over the premises in several ground coordinates, increasing the accuracy of the computer vision algorithms. Moreover, due to the inability to fly a drone in extreme weather conditions such as rain, strong winds or lightning, the system is limited to be used under few specifications. As swimming in extreme weather conditions is not preferred either, the system could be further improved to emit a warning signal if a person was to swim in any of the above weather conditions, bypassing the need to fly the drone.

Additionally, all the processing is done on the client-side of the applications on the Jetson Nano board, preventing any security and privacy issues that might arise due to the sensitive information inputted through the cameras.

For future developments convenience wise, the system could benefit by having an additional set of cameras to identify and verify a drowning or a hazardous activity on the premises. Accessibility could also be improved by extending the Android service to be an application both in Android and iOS platforms that could hold the details of each premise individually, making a centralized system that watches over the decentralized pool premises. Both drown and hazardous activity detection could be improved by gathering a nighttime dataset that increases the accuracy of the data in low light.


Fig.9. Drone residing prior to take-off

V. REFERENCES

[1] WHO, "Drowning," 27 April 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/drowning.

[2] CDC, "Drowning — United States, 2005–2009," CDC, 18 May 2012. [Online]. Available: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6119a4.htm. [Accessed 2021 October 2021].

[3] Q. Health, "The long-term impacts of non-fatal drowning," [Online]. Available: https://www.health.qld.gov.au/news-events/news/the-long-term-impacts-of-non-fatal-drowning. [Accessed 21 Feb 2020].

[4] L. Bugeja, "An analysis of stratagems to reduce drowning deaths of young children in private swimming pools and spas in Victoria, Australia," in *Int J Inj Contr Saf Promot*, 2013.

[5] A. S. Joost Bierensa, "Drowning in swimming pools," in *Microchemical Journal*, Rome, 2013.

[6] Local10.com, "What factors influence drowning risk?," local10.com, 13 October 2014. [Online]. Available: https://www.local10.com/family/2014/10/13/what-factors-influence-drowning-risk. [Accessed 21 January 2021].

[7] D. Hartmann, "Drowning and Beach-Safety Management (BSM) along the Mediterranean Beaches of Israel: A Long-Term Perspective," in *Journal of Coastal Research*, Beer-Sheva, 2006.

[8] AngelEye, "Safeguard," AngelEye.tech, [Online]. Available: https://www.angeleye.tech/us/us-safeguard/. [Accessed 25 November 2020].

[9] C. Zhang, "A Novel Camera-Based Drowning Detection Algorithm," in *IGTA 2015: Advances in Image and Graphics Technologies pp 224-233*, Berlin, 2015.

[10] A. I. N. Alshbatat, "Automated Vision-based Surveillance System to Detect Drowning Incidents in Swimming Pools.," in *In Proceedings of the 2020 Advances in Science and Engineering Technology International Conferences (ASET)*, Dubai, UAE, 2020.

[11] SwimEye, "A drowning detection system for any pool," Swimeye.com, [Online]. Available: https://swimeye.com/swimeye/. [Accessed 25 November 2020].

[12] L. Wenmiao, "A vision-based approach to early detection of drowning incidents in swimming pools," in *IEEE transactions on circuits and systems for video technology 14.2 159-178.*, 2004.

[13] E. Golan, "Drowning Detection gets real with Coral Manta 300," https://www.technology-innovators.com, [Online]. Available: https://www.technology-innovators.com/coral-drowning-detection-systems/. [Accessed 25 November 2020].

[14] Poseidon, "The Drowning Detection system," Poseidon, [Online]. Available: https://poseidon-tech.com/. [Accessed 25 November 2020].

[15] H. M. A., "The Swimmers Motion Detection Using Improved VIBE Algorithm," in *International Conference on Robotics and Automation in Industry (ICRAI)*, 2019.

[16] R. Hussin, "Digital image processing techniques for object detection from complex background image," in *Procedia Engineering 41 340-344.*, 2012.

[17] R. Girshick, "Fast R-CNN," in *IEEE international conference on computer vision*, 2015.

[18] S. Ren, "Faster R-CNN: Towards Real-Time Object," 2016.

[19] A. F. Joseph Redmon, "YOLOv3: An Incremental Improvement," in *Computer Vision and Pattern Recognition*, 2018.

[20] J. Hui, "Object detection: speed and accuracy comparison (Faster R-CNN, R-FCN, SSD, FPN, RetinaNet and YOLOv3)," Medium, 28 March 2018. [Online]. Available: https://jonathan-hui.medium.com/object-detection-speed-and-accuracy-comparison-faster-r-cnn-r-fcn-ssd-and-yolo-5425656ae359. [Accessed 12 10 2021].

[21] B. D. Prakash, "Near-drowning Early Prediction Technique Using Novel Equations (NEPTUNE) for Swimming Pools," IAG Firemark Singapore.

[22] l. V. Fares Jalled, "Object detection using image processing," in *arXiv preprint arXiv:1611.07791*, 2016.

[23] E. How-Lung, "An automatic drowning detection surveillance system for challenging outdoor pool environments".

[24] Z. M. Wenjie Zhan, "From Body Parts Responses to Underwater Human Detection: A Deep Learning Approach," 2020.

[25] A. Carballo-Fazanes, "The Visible Behaviour of Drowning Persons: A Pilot Observational Study Using Analytic Software and a Nominal Group Technique," *International Journal of Environmental Research and Public Health,* p. 17, 2020.

[26] A. Gopalakrishnan, "A neural temporal model for human motion prediction," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

[27] J. Redmon, "You Only Look Once: Unified, Real-Time Object Detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[28] R. Birch, "Review of a national water safety and drowning prevention model three years on: a sri lankan case study," in *Safety 2018 abstracts*, 2018.

[29] C. Yang, "Human-Pose-Estimation-Benchmarking-and-Action-Recognition," Github, 10 March 2019. [Online]. Available: https://github.com/ChengeYang/Human-Pose-Estimation-Benchmarking-and-Action-Recognition/blob/master/doc/command_benchmarking.md. [Accessed 10 10 2021].

[30] Z. Cao, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," in *Computer Vision and Pattern Recognition*, 2019.

[31] G. Hidalgo, "OpenPose: Whole-Body Pose Estimation," 2019.

[32] Google Play, "HelpStork," Google, 26 August 2021. [Online]. Available: https://play.google.com/store/apps/details?id=com.automatedlifeguard.push_notifs. [Accessed 12 10 2021].

[33] A. Dhiman, "Object Tracking using DeepSORT in TensorFlow 2," Medium, 27 October 2020. [Online]. Available: https://medium.com/analytics-vidhya/object-tracking-using-deepsort-in-tensorflow-2-ec013a2eeb4f. [Accessed 17 August 2021].

[34] WHO, "Drowning," WHO, 3 February 2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/drowning. [Accessed 17 January 2021].