Leveraging Deep Learning for Drowning and Swimming Prevention

TÜRKALP TÜRKER ÜNLÜ, ABAKAR LİMANE MAHAMAT, METİN TURAN
Department of Computer Engineering,
Istanbul Ticaret University,
Küçükyali, Inönü Cd. No. 4, 34854, Maltepe, Istanbul,
TURKEY

Abstract: - Drowning is a severe public health problem that has claimed many lives in Turkey. Our study focuses on a novel strategy to address this problem. We developed a semi-automated technique that uses drone technology and machine learning to prevent drowning in real time. Advanced convolutional neural network (CNN) models, including Xception, ResNet-50, and YOLOv8, were used in our approach. These various models were trained using a special dataset that included simulated drowning situations in the Turkish Aegean Sea and a collection of online images. Though fully automated operation cannot be ensured, this method greatly improves water safety by guiding the drone to the drowning event and alerting the relevant staff. The models performed admirably with relative accuracy rates of 82.1%, 83.40%, and 85.8%. This cutting-edge approach demonstrates how machine learning could fundamentally alter the way major health concerns are handled. It also demonstrates how, when integrated with traditional safety procedures and human supervision, technology can improve and assist human efforts to safeguard the public's health.

Key-Words: - Drowning prevention, machine learning, convolutional neural networks (CNN), YOLOv8, ResNet50, Xception, semi-automatic systems, public health.

Received: May 13, 2024. Revised: December 4, 2024. Accepted: January 7, 2025. Published: February 3, 2025.

1 Introduction

Drowning has emerged as a major public health concern in Turkey. Over 230,000 fatal drownings were recorded in 2019, making it a worldwide issue, primarily in countries with low and middle incomes, however it is especially severe in Turkey. An average of 935 accidental drownings per year took place in Turkey between 2010 and 2020. The mortality rate was 0.79 deaths per 100,000 inhabitants, with 66.1% of them leading to death. It is noteworthy that the highest prevalence of unintentional drownings was found in Bartın Province, which is situated along the southern Black Sea coast, [1].

In response to these alarming statistics, we have compiled a dataset from the Aegean Sea and the internet. The main objective of this dataset is to predict future drowning and swimming incidents using data from swimming and drowning incidents in the Aegean Sea, Turkey.

Research carried out in 2021 by [2] which employed deep learning algorithms to address the problem of deaths reported in swimming pools. Traditionally trained for event classification, high-precision deep neural networks (DNNs) like ResNet50, VGG16, and MobileNet were employed. Similarly, the study by [3] utilized an altered convolutional neural network (CNN) in conjunction with the YOLO technique for object recognition to design a deep learning-based underwater tracking system. In a record time frame of at least one minute, their system has demonstrated the ability to recognize and respond to a drowning instance in a swimming pool [4] by improving the YOLOv3 algorithm to solve the problem of finding and saving people in maritime mishaps, attaining a 72.17% detection accuracy for human targets at sea. [5] looked into the worldwide problem of drowning, which is one of the main causes of death for kids between the ages of 1 and 14 in swimming pools.

Convolutional Neural Networks (CNNs) such as SqueezeNet, GoogleNet, AlexNet, ShuffleNet, and ResNet50 were among the ones researchers used. ResNet50 showed the strongest prediction accuracy among them.

[6] has employed deep-learning techniques to detect drowning swimmers. Their approach involves an artificial intelligence-based motion recognition system that utilizes video processing technology and the OpenPose algorithm. Cameras installed at the bottom of swimming pools capture the spatial distribution of swimmers. By identifying key joint points in the human skeleton using OpenPose, the system determines whether a swimmer is in distress.

This real-time detection technique makes an important contribution to a decrease of drowning incidents in public pools. [7] uses 5G and deep learning technologies in their study to protect children from drowning in swimming pools. Using deep learning and image processing, they present a novel 5G-enabled drowning avoidance system that can identify inattentive parents or guardians in real time and notify them to concentrate on child monitoring. For this, the study used the ResNet-50 model.

Even though swimming pools have received a lot of attention from researchers, relatively few have focused on occurrences at sea. This study covers the gap by focusing on maritime disasters, especially in Turkey, where accidental drownings are common. We present a novel, semi-automated approach that uses drones and machine learning to avoid drowning events. Machine learning and drone technologies are used in our semi-automatic drowning avoidance system. By alerting the right people and guiding the drone to the potential drowning victim's location, the technology significantly enhances water safety procedures, even though total automation is not assured. Our methodology demonstrates the potential for change of modern facilities and machine learning techniques in grasping and adapting to the intricate nature of maritime environments. By significantly enhancing water safety and potentially saving lives, this approach demonstrates how machine learning may be used to address significant public health issues.

The sea presents a highly challenging environment due to its natural dynamics, such as waves and bubbles, as well as varying weather conditions (sunny or cloudy) that alter the sea’s color from dark to light. Additionally, a number of circumstances and elements, such as swimmer position, image contrast, and light levels, have a big impact on the learning process. To address these issues, we have created a dataset that includes a variety of these situations, and we are assessing how well contemporary deep learning methods work to address this issue. It has been demonstrated that convolutional neural networks, or CNNs, are an essential component of computer vision. When these algorithms are used, input images are processed first, after which they distinguish and assign weights to various attributes. Furthermore, the use of transfer learning, a complex deep learning method, has been very helpful in the advancement of this kind of model. A pre-trained model, such as MobileNet, DenseNet or Inception, is subsequently reused in this method as the foundation for a related task. This method is essential for saving time and computing resources, [8], [9], [10], [11], [12].

Initially, we used CNN without transfer learning as part of our study. Because of the previous challenges, the test results were not enough to reliably differentiate between images of swimming and drowning. As a result, the CNN model required a sizable, well prepared dataset. We used Xception, YOLOv8, and ResNet50 models with transfer learning to address this, and the results were satisfactory.

To deal effectively with the intricacies of marine environments, our research focuses on advanced transfer learning applications, in particular Xception, YOLOv8 and ResNet50. These conditions pose certain challenges to machine learning due to their intrinsic dynamism and weather- driven variations. To address these challenges, we present an extensive database covering a range of circumstances, such as swimmer positions, wave shapes, bubble formations and different light levels. In the process, we implement the CNN architecture known for its extraordinary efficiency on the ImageNet dataset, the Xception model. The CNN model has enabled humans to actively predict outcomes in situations such as swimming and drowning. Our advanced machine learning techniques for assessing and responding to the many characteristics of marine ecosystems are improving dramatically thanks to our attention to this advanced methodology.

2 Material and Methods

2.1 Data Preparation

The implementation of convolutional neural network (CNN) models found to be crucial for the classification of swimming and drowning incidents. 387 drowning and 376 swimming images make up the dataset used to train our model.

Depending on environmental conditions such as brightness, background, object distance, and camera characteristics, CNNs yield varying results. Consequently, we did not apply any special conditions or structures for image acquisition and we worked with images of varying resolutions.

The dataset is divided into two subsets: the training set and the test set. The test set comprises 20% of the total data, while the remaining 80% is used for training. This approach allows for a comprehensive evaluation of the model’s performance.

2.2 Deep Learning Models

Deep learning is fundamentally based on artificial neural networks and represents a sophisticated machine learning technique. It is capable of identifying complex features and relationships within large datasets by employing multilayer neural networks.

Numerous examples in the literature illustrate its application in the detection of health issues. For instance, [13] utilized artificial neural networks for the early detection of breast cancer.

We examined one pure CNN and three transfer-learning (pre-trained) CNN models: YOLOv8, ResNet50, and Xception.

2.2.1 CNN Architecture

Convolutional neural networks (CNNs) are a potent class of deep learning algorithms that are frequently used for computer vision tasks such as segmentation, object identification, and picture categorization. CNNs directly extract features from unprocessed visual data, in contrast to conventional feature extraction techniques, [14]. Hierarchical feature extraction is made possible by CNNs' use of convolutional layers to apply filters to input images. These attributes include generalizations like shapes and objects as well as low-level elements like edges and textures. Many layers, including convolutional, pooling, and fully linked layers, make up a typical CNN architecture (Figure 1).

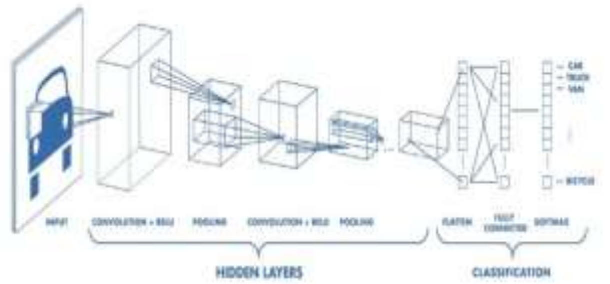

Fig. 1: CNN architecture, [15]

In the present research, we employed models that include transfer learning, such as Xception, YOLOv8, and ResNet50, as well as CNN with no transfer learning. Applying previously trained models to various datasets and feature domains is known as transfer learning. This method uses a model that was once created for a particular task as the basis for another. Transfer learning's main advantages is its capacity to increase learning effectiveness and model performance in the intended task, especially in situations when labeled data is limited.

2.2.2 YOLOv8 Architecture

The first YOLO model emerged in 2015. The YOLOv8 model used in this study was introduced in 2023 (Figure 2).

Fig. 2: Timeline of YOLO models, [16]

Examples in the literature demonstrate the application of deep learning models in the detection of health issues. For instance, [17] utilized the YOLO model for the detection of skin cancer.

Key factors contributing to health include cleanliness, hygiene, and a healthy diet. [18] employed the YOLO algorithm to detect overflowing garbage cans. [19] utilized the YOLO algorithm to detect spoiled fruits.

YOLOv8 is a sophisticated object detection model that uses a fully convolutional neural network, which consists of two primary parts: a head for class prediction and a backbone for feature extraction. The anchorless detection method and the new convolutional layer C2f, which is intended to improve feature detection efficiency, are noteworthy additions (Figure 3).

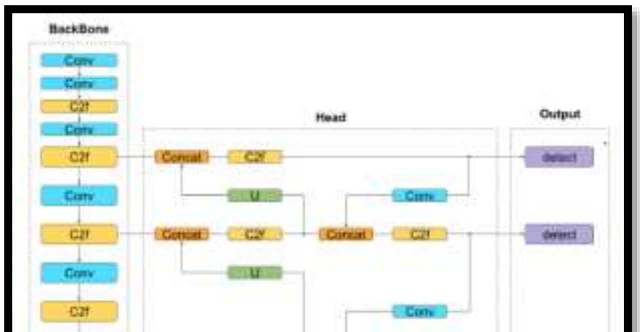

Fig. 3: The visualization of YOLOv8's architecture, where arrows denote the flow of data between layers, [20]

Its superior speed and accuracy make it highly adaptable for a wide range of applications and hardware platforms, [20], [21].

YOLOv8, the most recent release of the YOLO family, has demonstrated that its three main architectural components: the head, neck, and backbone allow it to surpass previous models by more accurately detecting objects at different scales and by progressively reducing the computing load.

Visual features are extracted from the input image using the Backbone. This kind of YOLOv8 model uses Partial Cross-Connections (CSP) and a Path Aggregation Network (PANet) to enhance the Convolutional Neural Network (CNN). The YOLOv8 architecture's rapid image processing ensures that the image quality will always be nearly constant, [22].

Visual features are extracted from the input image using the Backbone. This kind of YOLOv8 model uses Partial Cross-Connections (CSP) and a Path Aggregation Network (PANet) to enhance the Convolutional Neural Network (CNN). The YOLOv8 architecture's rapid image processing ensures that the image quality will always be nearly constant, [23].

The Neck of YOLOv8 plays an important role in refining and fusing the features extracted by the Backbone. To incorporate features from different layers, the combination of Feature Pyramid Networks (FPNs) and Path Aggregation Networks (PANets) is crucial in this type of architecture. These strategies make it easier to detect objects of different sizes, which makes it crucial to assimilate particularity at various sizes in order to accurately identify small and large objects in a single image.

Finally, the last couche, which is the head of the architecture YOLOv8, or terminal component, has produced the final predictions. Additionally, it enables the generation of englobant boxes, class probabilities, and confidence scores for each object found. With the support of this anchor-free detection technique, YOLOv8 normalizes the model and gradually decreases the number of parameters to speed up inference times, [24].

2.2.3 ResNet50 Architecture

In order to address this issue, ResNet was introduced in 2016 with the goal of improving the accuracy and speed of neural networks with unique characteristics.

The ResNet technology's architecture consists of 50 couches, as seen in Figure 4. She was created to promote selective neuron activation while prioritizing the acquisition of new features over the relearning of existing ones, hence optimizing the model's learning process.

All of the layers mentioned above are presented in the 2016 paper "Deep Residual Learning for Image Recognition." The vanishing gradient problem is the main obstacle that arises when deep neural networks are trained by incorporating shortcut connections that avoid one or more layers, [25]. This issue is alleviated in part by this architecture. Convolutional layers can be incorporated through these many levels, which enable them to be grouped and completely connected. It introduces a significant innovation: the use of residual blocks, in which a layer's input is added directly to a subsequent layer's output. Training very deep networks is made easier by this "residual" connection, which enables the network to learn belonging mappings without experiencing any performance deterioration.

In computer vision, ResNet 50 was one of the main models, which integrated into itself several varieties of designs and was capable of performing tasks such as object identification, image segmentation and image classification. . It has demonstrated its exceptional effectiveness when training on massive datasets, such as ImageNet, while maintaining computational efficiency and excellent accuracy.

This architecture is more well-known for its wide range of deep learning applications. For instance, ResNet50 has been used to recognize emotions in the research of [26]. Additionally, she was used in earlier studies to identify a variety of health issues. For instance, the study by [27] also used this architecture to automatically predict the COVID-19 based on thoracic images. In order to facilitate early detection of skin cancer, the author of the paper [28] used this design for the classification of skin lesions. Additionally, research conducted by [29] has demonstrated its effectiveness in the early detection and classification of cancer in the body. Furthermore, ResNet50 is effectively employed in diagnosing eye diseases. For instance, it excels in detecting diabetic retinopathy by analyzing retinal images, demonstrating its strong diagnostic accuracy, [30].

Fig. 4: Structure of the ResNet50 model, [5]

2.2.4 Xception Architecture

The deep learning network used was the Xception model, developed by Google and renowned for its depth-separable convolutions (DSC). DSC divides the computation into depthwise and pointwise convolutions to improve efficiency. Xception, a variant of Inception, scores better on ImageNet than Inception V3 due to its improved usage of parameters. It comprises 36 convolutional layers organized into 14 modules, forming the foundation for feature extraction. For image classification, a logistic regression layer follows this base. Each module integrates batch normalization, ReLU activation, and a 3x3 depth-separable convolution kernel. Data traverses through input, middle, and output streams (Figure 5). With the exception of the first and last modules, all modules incorporate linear residual connections. Batch normalization normalizes the eigenvalues, thereby balancing the output distribution of each layer. The ReLU activation function introduces non-linearity, enhancing the model's adaptability, [31].

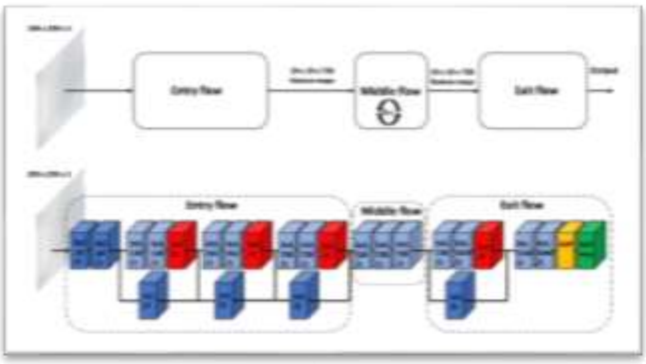

Fig. 5: Xception CNN structure for detection and classification tasks, [32]

2.3 Processing

Preprocessing the input images is essential for enhancing the model's accuracy, preventing overfitting, and improving its generalization capability. Initially, we resize all images. Subsequently, we normalized the pixel values by dividing them by the maximum pixel values of the captured images. Following this, we applied data augmentation techniques, including random translation, random zoom, random shear, random flip, random brightness, and random rotation, to artificially increase the number of images used in model development.

The parameters used for data augmentation are detailed in Table 1.

We experimented with various structures and configurations, including 2D convolution, global averaging, dropout, and batch normalization. To prevent overfitting and conserve computational resources, we also incorporated an Early-Stopping mechanism, which halts training when the model's performance on the validation set ceases to improve.

Table 1. Data augmentation parameters and their associated values

Data Augmentation Parameters	***Value***
Random Translation (height_shift_range)	*0.2*
Random Translation (width_shift_range)	*0.2*
Shear Intensity (in radian)	*0.2*
Random Zoom	*0.2*
Random Flip	*True*
brightness_range	*[0.2, 1.0]*
rotation_range	*40*

3 Results and Discussion

3.1 Evaluation Measures

Machine learning models are evaluated using several performance measures, including accuracy, sensitivity, specificity, precision, and the F1 score. Precision measures the proportion of correctly identified instances in the test dataset. Sensitivity, or recall, quantifies the successful prediction of drowning cases relative to all drowning instances in the dataset. Precision evaluates the proportion of accurately predicted drowning cases among all predicted drowning cases. Specificity calculates the proportion of non-drowning instances accurately identified among all non-drowning instances in the dataset. These measures provide a comprehensive assessment of model performance.

The computation of these metrics can be accomplished through the application of the corresponding mathematical equations. These equations serve as the foundation for quantifying the performance of machine learning models.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FN + FP)} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

$$F1 - Score = \frac{2(Precision \times Recall)}{Precision + Recall} \tag{5}$$

3.2 Model Training and Validation

As illustrated in Figure 6, the CNN model without transfer learning was trained for 40 epochs on a dataset comprising 763 images from the Aegean Sea, utilizing an 80-20 split for training and testing. The model’s precision began at 50% and increased to 66% by epoch 40, indicating insufficient learning and generalization.

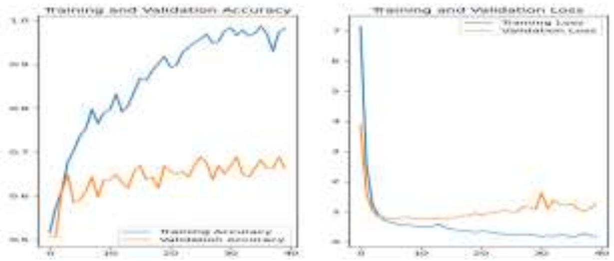

Fig 6: Performance results of the CNN model without transfer learning

The YOLOv8 model was trained across 10 epochs using a dataset of 763 photos, with an 80-20 split for training and testing, as shown in Figure 7. Starting at 64%, the model's accuracy peaked at the fifth epoch and ended at 82.1% at the tenth. This pattern demonstrates how well the model can learn and generalize.

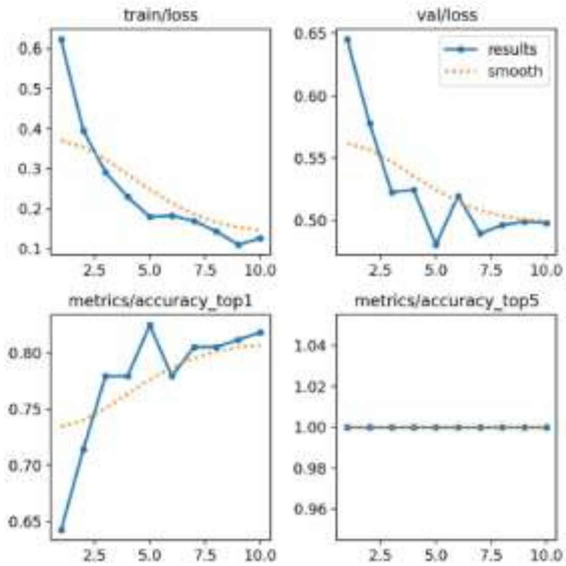

Fig. 7: Performance results of YOLOv8 model

Figure 8 presents the accuracy progression of a ResNet50 model, designed using MATLAB, over 40 epochs. The model was trained on a dataset of 763 images (387 drownings and 376 swimming), with an 80-20 split for training and testing. The validation accuracy started at 82% and gradually increased to 83.40% by the 40th epoch, indicating the model’s effective generalization to unseen data. From all the above, the training of our model was satisfactory because the model started its training with an accuracy of 98% from the 40th epoch and saw a slight increase of 1% that gives an accuracy of 99% at the end of the training. The essential information on the learning dynamics and the adaptation of the capacity to biases and variance has been proven by this model, these two elements of which show how often the model is very efficient during prediction.

Figure 8 illustrates how the loss reaches roughly 0.2 in certain epochs before stabilizing.
The significance of selecting the ideal number of epochs for model learning is highlighted in this figure.

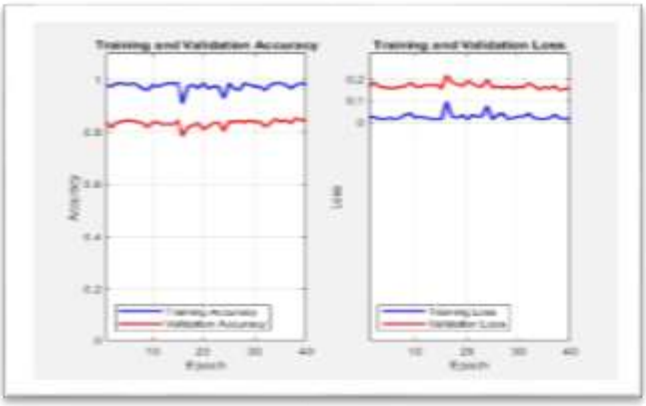

Fig. 8: Performance results of ResNet50

In order to train the Xception model over 18 epochs, as illustrated in Figure 9, we employed a dataset of 763 images, which was split into 80% for training and 20% for testing. The model's initial accuracy peaked at the 15th epoch at 64%. By providing a precision that comparatively increased by 85.8% at the 18th epoch at the end of the training, the training demonstrated successful generalization. This adequately demonstrates how often the model is effective even at the most basic level. At first, the precision was 64% and peaked at the 15th epoch, demonstrating the optimal training point. At the 18th epoch, the accuracy stabilized at 85.8%, indicating strong generalization to unknown data. The design of the Xception model, with its depth wise separable convolutions, was found to be quite effective in reducing parameters and processing costs; it only needs to increase the accuracy. This development demonstrates the model's learning efficiency and flexibility, making it a reliable means of making accurate predictions in practical situations.

Figure 10 illustrates how the Xception model may learn particular characteristics, reduce losses, and produce good prediction performance by displaying increases in the model loss. This setup highlights the importance of choosing the optimal number of epochs for optimal model performance.

To maintain the rationality of the equalities, all layers were frozen at the start of training. After unfreezing the final 20 layers, the model will be trained on new data. To enhance the model's learning capabilities, we subsequently included three more blocks: convolution, batch normalization, max pooling, and dropout layers. The next steps were completely connected layers and global average pooling. This methodology successfully balanced the use of previously learnt data with the collection of fresh data, optimizing the model's performance. Applying freezing and unfreezing layers selectively has been shown to be very effective in preserving significant pre-trained features while adapting to new input, ensuring a comprehensive learning process. In order to achieve high accuracy and generalization in prediction tasks, it has been necessary to strike a balance between using new data and existing knowledge.

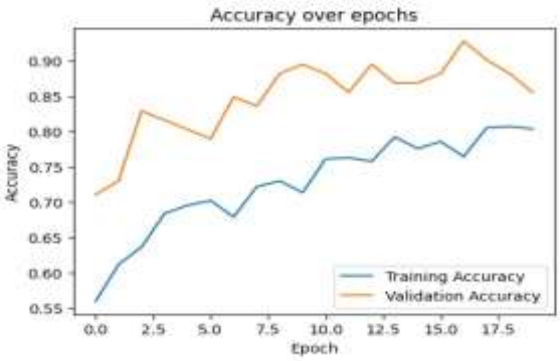

Fig. 9: Accuracy graph of Xception model

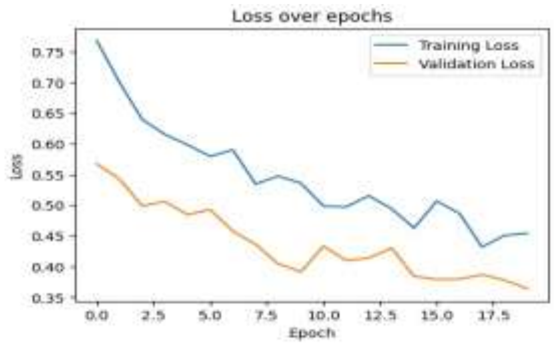

Fig. 10: Loss graph of the Xception model

The appropriate output shapes, the sizes of the 96x96 images under study in the various levels of our model, and the total number of parameters are all shown in Table 2. Naturally the roughly 23,485,482 parameters in the model, 5,452.8 are untrainable. It is important to note that layers such as Max Pooling, Dropout, and Global Average Pooling do not contribute to the trainable parameters, as they inherently lack them.

The Xception model mitigates overfitting through the use of dropout layers, which randomly disconnect a connection with a 50% probability during testing. This approach, while doubling convergence time, prevents overfitting. The model is capable of classifying millions of images into various object categories, providing labels, and determining probabilities for each category.

Detailed specifications of the Xception CNN architecture are shown in Table 2.

Table 2. Detailed characteristics of the modified Xception CNN architecture applied to resize input images

Layer (Type)	***Output Shape(Input Size = 96×96)***
Xception Block	*(None, 96, 96, 3)*
Convolution 2D	*(None, 47, 47, 32)*
Batch Normalization	*(None, 47, 47, 32)*
Activation	*(None, 47, 47, 32)*
Convolution2D	*(None, 45, 45, 64)*
Convolution 2D	*(None, 23, 23, 128)*
Batch Normalization	*(None, 23, 23, 128)*
Max Pooling 2D	*(None, 23, 23, 128)*
Dropout	*(None, 512)*
Convolution 2D	*(None, 3, 3, 2048)*
Batch Normalization	*(None, 23, 23, 128)*
Max Pooling 2D	*(None, 23, 23, 128)*
Dropout	*(None, 512)*
Dense	*(None, 1024)*
Dense	*(None, 1024)*
Dense	*(None, 1024)*
Global Average Pooling 2D	*(None, 2048)*
Dense	*(None, 2)*
Total Parameters	*23485482 (89.59 MB)*
Trainable Parameters	*23430954 (89.38 MB)*
Non-Trainable Parameters	*54528 (213.00 KB)*

3.3 Assessment and Validation of the Model

The confusion matrices shown in Figure 12, Figure 13, Figure 14 and Figure 15 provide a summary of the results of the evaluation of each model's accuracy applying the test dataset (Figure 11). Table 3 presents a summary of the test results for each model, which are described in this thorough study.

With a comparatively low accuracy of 0.67, the CNN without transfer learning outperformed the drowning model in swimming. With an accuracy of 0.88, ResNet50 outperformed the other models in the evaluation of transfer learning performance, especially in swimming. With an accuracy of 0.86, Xception trailed closely behind and demonstrated impressive swimming results. In contrast, YOLOv8 performed somewhat better in swimming, achieving an accuracy of 0.83 in swimming and 0.82 in drowning. These findings show that ResNet50 and Xception are especially good at tasks involving swimming, even if all models perform well overall.


Fig. 11: Some examples of predicted images

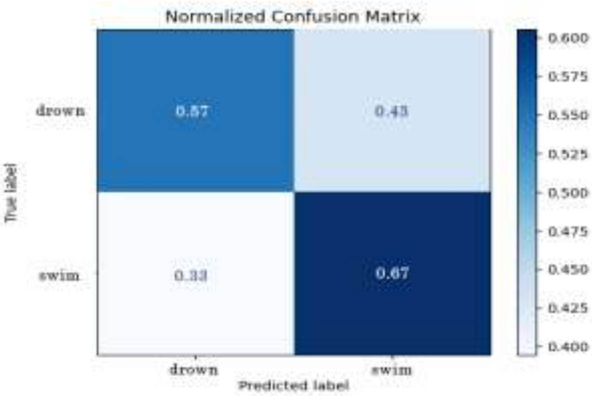

Fig. 12: CNN confusion matrix

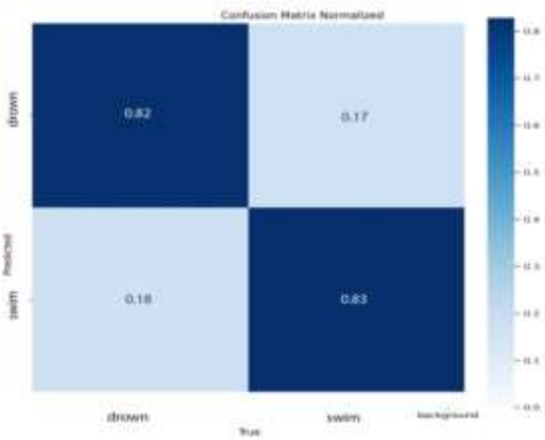

Fig. 13: YOLOv8 confusion matrix

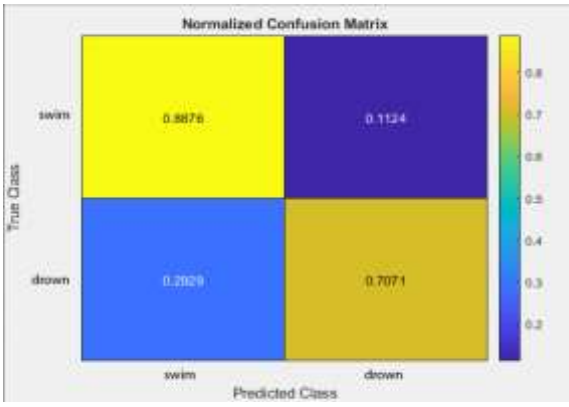

Fig. 14: ResNet50 confusion matrix

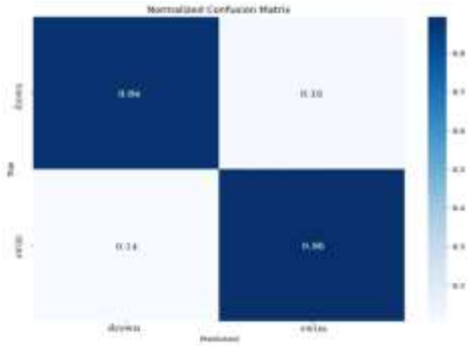
Fig. 15: Xception confusion matrix

The best accuracies of the implemented pure CNN model and CNN models with transfer learning (YOLOv8, ResNet50, Xception) are given in Table 3.

Table 3. Best outcomes of the models

CNN Model	*Validation Accuracy*	*# of Epochs*	*Learning Rate*
CNN	0.66	40	1.00E-04
YOLOv8	0.821	10	1.00E-04
ResNeT50	0.834	40	1.00E-04
Xception	0.858	18	1.00E-04

4 Conclusion and Future Work

Considering a particular focus on Turkey, this study concludes by offering a revolutionary solution to the serious public health problem of drowning. Our proposal is a cutting-edge, semi-automated solution that uses drones and machine learning to avoid drowning events in real time. Our method uses sophisticated Convolutional Neural Networks (CNNs), including Xception, ResNet50, and YOLOv8, and integrates transfer learning to improve the models' performance.

Our models exhibit remarkable adaptability to real-world circumstances, having been trained on a rich dataset that includes 387 drownings and 376 swimming photos. Convolutional neural networks, or CNNs, improve the model's practical applicability by analyzing photos taken in natural settings. However, because of the complexity of the water and the similarity of the acts (drowning and swimming), the CNN model without transfer learning failed. For the model to understand the distinctions and make a clear distinction between these activities, more instances are needed.

Conversely, the transfer learning models (YOLOv8, ResNet50, Xception) demonstrated superior performance due to their prior training on extensive datasets. These models achieved their highest accuracies at different epochs, with the training process involving multiple iterations to minimize error and optimize weights. Consequently, YOLOv8, ResNet50, and Xception attained accuracies of 82.1%, 83.40%, and 85.8%, respectively. This supports our allegation, although it would not ever reach 100% success in the realistic sea environment.

Most of the research in the literature, such as [3] and [5], applied to a swimming pool environment which is mostly clear and determined. However, we focused on the sea environment which has indeterministic parameters mostly.

In general, our methodology highlights the revolutionary possibilities of innovative machine learning methods in comprehending and responding to the intricate features that make up aquatic environments. The promise for this technology to improve water safety and save lives is vital, highlighting the effectiveness of machine learning in resolving important public health concerns.

Future research should prioritize expanding the dataset to encompass more diverse scenarios and larger sizes to enhance model performance and adaptability. Investigating other recently developed CNN architectures could potentially yield superior results. Real-time system efficiency enhancement in dynamic and real-world environments is another useful technique. Furthermore, modern technology must be included into both new and existing protective systems, such as automated lifeguards and alert systems. Lifeguards and other safety personnel would find the system easier to use if it had an intuitive interface. Defining guidelines for its suitable application and taking into account the ethical and privacy consequences of this technological surveillance are just as important. Large-scale horizontal study could yield information on the system's long-term efficacy and its effect on drowning rates. Lastly, using transfer learning to additional datasets that deal with health or water safety concerns may demonstrate the scalability of this approach.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

The authors wrote, reviewed and edited the content as needed and they have not utilized artificial intelligence (AI) tools. The authors take full responsibility for the content of the publication.

References:

[1] S. Alkan and U. Karadurmuş, "Risk assessment of hazard factors on drowning incidents in Turkey", *Natural Hazards*, 2023, vol. 118, pp. 2459-2475. doi: 10.1007/s11069-023-06095-7.

[2] S. Hasan, J. Joy, F. Ahsan, H. Khambaty, M. Agarwal, and J. Mounsef, "A Water Behavior Dataset for an Image-Based Drowning Solution," in *2021 IEEE Green Energy and Smart Systems Conference (IGESSC)*, Long Beach, CA, USA, Nov. 2021, pp. 1–5. doi: 10.1109/IGESSC53124.2021.9618700.

[3] U. Handalage, N. Nikapotha, C. Subasinghe, T. Prasanga, T. Thilakarthna, and D. Kasthurirathna, "Computer Vision Enabled Drowning Detection System," in *2021 3rd International Conference on Advancements in Computing (ICAC)*, Colombo, Sri Lanka, Dec. 2021, pp. 240–245. doi: 10.1109/ICAC54203.2021.9671126.

[4] D. Li, L. Yu, W. Jin, R. Zhang, J. Feng, and N. Fu, "An Improved Detection Method of Human Target at Sea Based on Yolov3," in *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, Guangzhou, China, Jan. 2021, pp. 100–103. doi: 10.1109/ICCECE51280.2021.9342056.

[5] M. Shatnawi, F. Albreiki, A. Alkhoori, and M. Alhebshi, "Deep Learning and Vision-Based Early Drowning Detection," *Information*, vol. 14, no. 1, p. 52, Jan. 2023, doi: 10.3390/info14010052.

[6] J. -X. Jian and C. -M. Wang, "Deep Learning Used to Recognition Swimmers Drowning," *2021 IEEE/ACIS 22nd International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Taichung, Taiwan, 2021, pp. 111-114, doi: 10.1109/SNPD51163.2021.9704884.

[7] J.C. Cepeda-Pacheco and M.C. Domingo "Deep Learning and 5G and Beyond for Child Drowning Prevention in Swimming Pools", *Sensors*, vol. 22, no. 19, 2022, 7684. doi: 10.3390/s22197684

[8] L. Moataz, G. I. Salama, and M. H. AbdElazeem, "Skin Cancer Diseases Classification using Deep Convolutional Neural Network with Transfer Learning Model," *J. Phys.: Conf. Ser.*, vol. 2128, no. 1, p. 012013, Dec. 2021, doi: 10.1088/1742-6596/2128/1/012013.

[9] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 16, 2017, *arXiv*: arXiv:1704.04861, [Online]. http://arxiv.org/abs/1704.04861 (Accessed: July 19, 2024).

[10] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, Jul. 2017, pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.

[12] F. Zhuanh, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A Comprehensive Survey on Transfer Learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.

[13] Z. Sultana, Md. A. R. Khan, and N. Jahan, "Early Breast Cancer Detection Utilizing Artificial Neural Network," *WSEAS Transactions on Biology and Biomedicine*, vol. 18, pp. 32-42, 2021, https://doi.org/10.37394/23208.2021.18.4.

[14] J. Gupta, S. Pathak, and G. Kumar, "Deep Learning (CNN) and Transfer Learning: A Review," *Journal of Physics: Conference Series*, vol. 2273, p. 012029, 2022. doi:10.1088/1742-6596/2273/1/012029.

[15] The Mathworks, Introducing Deep Learning with MATLAB, 2018, [Online]. https://www.mathworks.com/campaigns/offers/deep-learning-with-matlab.html (Accessed: July 19, 2024).

[16] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, 2023, pp. 1680–1716. doi:10.3390/make5040083.

[17] N. Aishwarya, M. Prabhakaran K, F. T. Debebe, M. S. S. Reddy, and P. Pranavee, "Skin Cancer diagnosis with Yolo Deep Neural Network," *Procedia Computer Science*, vol. 220, 2023, pp. 651-658. doi:10.1016/j.procs.2023.03.083.

[18] M. Panmuang and C. Rodmorn, "Garbage Detection using YOLO Algorithm for Urban Management in Bangkok," *WSEAS Transactions on Computer Research*, vol. 12, pp. 236-243, 2024, https://doi.org/10.37394/232018.2024.12.23.

[19] P. Kanupuru and N. V. Uma Reddy, "A Deep Learning Approach to Detect the Spoiled Fruits," *WSEAS Transactions on Computer Research*, vol. 10, pp. 74-87, 2022, https://doi.org/10.37394/232018.2022.10.10.

[20] M. Sohan, T. Sai Ram, and Ch. V. Rami Reddy, "A Review on YOLOv8 and Its Advancements", *Springer*, 2024, pp. 529–545. doi: 10.1007/978-981-99-7962-2_39.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[22] C. Y. Wang , H. Y. Mark Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, "CSPNet: A New Backbone that can Enhance Learning Capability of CNN", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, Jun. 2020, pp. 1571-1580. doi: 10.1109/CVPRW50498.2020.00203.

[23] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, Feature Pyramid Networks for Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Jul. 2017, pp. 936-944. https://doi.org/10.1109/CVPR.2017.106.

[24] H. Law and J. Deng, CornerNet: Detecting Objects as Paired Keypoints. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, Sep. 2018, 765-

781. https://doi.org/10.1007/978-3-030-01264-9_45.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90.

[26] [26] I. S. Ahmad, S. Zhang, S. Saminu, L. Wang, A. El K. Isselmou, Z. Cai, I. Javaid, S. Kamhi, and U. Kulsum, "Deep Learning Based on CNN for Emotion Recognition Using EEG Signal," *WSEAS Transactions on Signal Processing*, vol. 17, pp. 28-40, 2021, https://doi.org/10.37394/232014.2021.17.4.

[27] M. Elpeltagy and H. Sallam, "Automatic prediction of COVID-19 from chest images using modified ResNet50," *Multimedia Tools and Applications*, vol. 80, 2021, pp. 26451–26463. doi: 10.1007/s11042-021-10783-6.

[28] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, vol. 542, no. 7639, 2017, 115-118. https://doi.org/10.1038/nature21056.

[29] D. A. Ragab, M . Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines", *PeerJ,* Vol. 7, 2019, pp. 1-20. https://doi.org/10.7717/peerj.6201.

[30] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, and D. R. Webster, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs", *JAMA*, 316(22), 2016, 2402-2410. https://doi.org/10.1001/jama.2016.17216.

[31] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, Jul. 2017, pp. 1800–1807. doi: 10.1109/CVPR.2017.195.

[32] E. Westphal and H. Seitz, "A machine learning method for defect detection and visualization in selective laser sintering based on convolutional neural networks," *Additive Manufacturing*, vol. 41, p. 101965, May 2021, doi: 10.1016/j.addma.2021.101965.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

- Abakar Limane Mahamat and Türkalp Türker Ünlü carried out the implementation of techniques.
- Metin Turan prepared a data set and evaluated the results.

The Authors equally wrote and revised the manuscript.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This research was funded by the Istanbul Commerce Chamber as part of their social responsibility initiatives.

Conflict of Interest

The authors declare that they have no relevant conflicts of interest pertaining to the content of this article.