Performance Analysis of YOLO Architectures for Surgical Waste Detection in Post-COVID-19 Medical Waste Management

Zishan Ahmed
BSc CSE
American International University, Bangladesh
Dhaka, Bangladesh
20-42085-1@student.aiub.edu

Shakib Sadat Shanto
BSc CSE
American International University, Bangladesh
Dhaka, Bangladesh
20-43074-1@student.aiub.edu

Samin
BSc CSE
American International University, Bangladesh
Dhaka, Bangladesh
20-42408-1@student.aiub.edu

Abstract— In the wake of the COVID-19 outbreak, there has been a dramatic uptick in the need for efficient medical waste management, making it imperative that we develop more surgical waste management. Used surgical masks and gloves are examples of potentially infectious materials that are the subject of this research. By utilizing its real-time object detection capabilities, the You Only Look Once (YOLO) deep learning-based object detection algorithm is used to identify surgical waste. Using the MSG dataset, we undertake a deep dive into the performance of three different YOLO architectures (YOLOv5, YOLOv7, and YOLOv8). According to the findings, YOLOv5-s, YOLOv7-x, and YOLOv8-m all perform exceptionally well when it comes to identifying surgical waste. YOLOv8-m was the best model, with a mAP of 82.4%, among these three. To mitigate post-COVID-19 infection risks and improve waste management efficiency, these results can be used to the creation of automated systems for medical waste sorting.

Keywords— COVID-19, Surgical Waste, YOLO, Object detection, YOLOv8

I. Introduction

The post-COVID-19 era has seen an increase in the demand for safe and effective medical waste management, particularly in healthcare facilities. It has resulted in substantial modifications to the medical waste management system. These changes have made it necessary to implement a waste segregation procedure that is both more efficient and effective in order to stop the further spread of the virus. In the event that it is not disposed of in the appropriate manner, surgical waste, which includes old masks and gloves in addition to other potentially infectious substances, presents a high danger of infection. In order to prevent the spread of infectious diseases, one of the most important stages in this process is the appropriate identification and separation of hazardous medical waste, including surgical waste.

In this study, we have used MSG dataset [1] which contains 1153 images of Mask, Bio-hazard Symbol and Gloves. A deep learning-based object recognition method called You Only Look Once (YOLO) [2] has achieved outstanding results in a variety of applications, including the analysis of medical images [3] [4] [5]. Its success can be linked to the fact that it only requires one neural network to detect objects in real-time, making it an excellent option for identifying surgical waste.

However, a comprehensive performance study of several YOLO versions is necessary before selecting the most suitable YOLO architecture for post COVID-19 surgical waste detection. Accuracy, speed, and computing resources should all be taken into account in this research since they are very important in deciding if a YOLO architecture is suitable for detecting surgical waste. In order to detect post COVID-19 surgical waste, including used masks, gloves, and other biohazard materials, this study will do a thorough performance analysis of several YOLO architectures. The study we are conducting will compare the performance of anchor-based variants of YOLOv5 [6], YOLOv7 [7] and recently published anchor-free YOLOv8 variants utilizing a variety of metrics.

The results of this study will be helpful in the development of automated systems for the segregation of medical waste, which will lower the risk of infection in both the general public and among healthcare personnel. Additionally, this research will aid in the development of a medical waste management system that is more effective and dependable, particularly in the post-COVID-19 era.

II. Literature Review

Increased surgical waste, including used masks and gloves, which provide a significant risk of infection if improperly disposed of, has been brought on by the COVID-19 pandemic. It is vital to detect and separate such hazardous material in order to prevent the virus from spreading further.

A number of research have been carried out to categorize or find surgical waste. In their study, Chen *et al.* [8] assembled a video collection of four waste objects (gloves, hairnet, mask, and shoe cover) and suggested a motion detection-based technique to extract valuable frames from it. They suggested an architecture that included characteristics from 2D and 3D convolutional neural networks to categorize waste videos. On their dataset, their suggested approach had a 79.99% accuracy rate. Themistocleous *et al.* [9] used Sentinel-2 pictures from orbit to find floating plastic liters. Kumar *et al.* [10] suggest an AI-based system for classifying COVID-related medical waste. Prior to the commencement of the recycling process, the waste type classification was carried out using image texture-dependent features, which essentially assisted to give accurate sorting and classification. With an accuracy of 96.5%, SVM classifier performs best in their study. Ferdous and Ahsan [1] presents a method for identifying different kinds of infectious COVID waste. They used several YOLO architectural versions for their investigation. When compared, YOLOX performs better than the other architectures, with a mAP of 92.49%. Panwar *et al.* [11] used AquaVision, a deep learning-based detection algorithm, using the AquaTrash dataset. With a mean Average Precision (mAP) of 81%, their suggested model can identify and categorize the various pollutants and hazardous waste floating in the waters and along the coast. Mehendale *et al.* [12] aimed to create an automated computer vision system for medical waste separation that can identify and classify medical waste into four categories: cotton, cotton gloves, cotton masks, and cotton syringes. To achieve this, they developed a model using transfer learning on the AlexNet deep learning network. The system's training resulted in an accurate categorization of medical waste items, achieving a validation accuracy of 86.17%. In their study, Buragohain *et al.* [13] developed a COVID-19 waste detection model and focused on identifying syringes, masks, and hand gloves. To compare the accuracy of different models on their dataset, they trained several CNN models. Among them, the most precise model was EfficientDet D0, which achieved a mean average accuracy of 82%.

The literature review suggests YOLO frameworks can be efficient in identifying and categorizing surgical waste such as biohazardous materials, masks, and gloves after the COVID-19 pandemic. We experiment with various YOLO models on the MSG dataset, comparing their performance through training and testing. Moreover, we evaluate and differentiate the recently released YOLOv8 architecture from the relatively novel YOLOv7 and YOLOv5 model. Overall, our study focuses on analyzing the effectiveness of different YOLO architectures in detecting and sorting post-COVID-19 surgical waste.

III. PROPOSED METHOD

A. Framework

The process underlying object detection in the YOLO architecture is depicted in Figure 3. An input image is initially transmitted into the YOLO network, and distinctive features are extracted via the network's backbone. The extracted features are then utilized by the backbone network to generate a feature pyramid, which is then passed to the head network. The head network has two primary functions: regression of bounding frames and classification of objects. The output of the prediction phase may include any combination of the three desired categories: masks, gloves, and biohazard items.

In addition to the architecture, a novel dataset for detecting and managing infectious refuse in our environment has been curated. This dataset captures the diversity of real-world variations, angles, states, and textures. By incorporating such a wide variety of samples, the system's robustness and adaptability to numerous situations are improved.

B. Objective

This study's primary objective is to detect surgical debris and biohazard symbols accurately and in a reasonable amount of time. To achieve this objective, numerous YOLO architectures are analyzed, each of which serves a different purpose. In addition, two distinct varieties of YOLO models are chosen: one employing an anchor-based training mechanism and the other an anchor-free training mechanism.

Table-1 provides an exhaustive summary of the object detection models utilized in this investigation. Three anchor-based models and one anchor-free model are displayed in the table. Each model was meticulously selected based on its distinct qualities and capabilities.

We used the YOLOv5 architecture in its many forms, including the YOLOv5-s, YOLOv5-m, YOLOv5-l, and YOLOv5-x versions. There were two different designs utilized for YOLOv7: YOLO-v7 and YOLOv7-x. We used the YOLOv8-s, YOLOv8-m, YOLOv8-l, and YOLOv8-x architectures, all of which are somewhat different from one another. The letters "s," "m," "l," and "x" stand for "small," "medium," "large," and "extra-large," respectively. The hypothesis suggests that larger models would often have higher levels of accuracy than their smaller equivalents. In contrast, smaller models offer faster processing rates than their larger counterparts. Consequently, the decision between model size and performance depends on the specific perspective or application being developed. Therefore, it is necessary to conduct a thorough evaluation of all versions of YOLOv5, YOLOv7, and YOLOv8 in order to provide a comprehensive description of their performance.

In addition, it is important to evaluate the efficacy of anchor-based and anchor-free detectors. Understanding the distinctions and capacities of these detection mechanisms is essential for determining their suitability for particular applications.

Training Mechanism	Architecture
Anchor-based	YOLOV5s YOLOV5m YOLOV5l YOLOV5x YOLOV7 YOLOV7x
Anchor-free	YOLOV8s YOLOV8m

	YOLOV8l YOLOV8x

Table:1

C. *Training*

In our case, we trained using 80% of the data and validated with 20%. The whole thing was trained and validated using Google Cloud (Google Colaboratory). With an input image size of 416x416, the training procedure lasted for 40 iterations. In order to train the YOLOv5 architecture, we used the PyTorch environment and followed the training approach created by Ultralytics, a leading firm in the industry. Similarly, YOLOv7 and YOLOv8 were trained using pretrained weights and the same construction and procedures given by Ultralytics. For both models, we utilized a batch size of 16 and 30.

All models were trained in PyTorch environment and SGD optimizer was used. Table-2 represents the training hyper parameters.

Model	Learning Rate	Decay	Batch Size
YOLOv5	0.01	0.0005	16/30
YOLOv7	0.01	0.0005	16/30
YOLOv8	0.01	0.0005	16/30

Table:2

IV. Results

A. *Findings*

From table-3, we can see the results for all the YOLO models used in this study.

MODEL	Batch Size	Mask AP	Gloves AP	Biohazard AP	mAP
YOLOV5s	16	94.5	80.8	63.8	79.7
	30	93.7	81.8	69.9	81.8
YOLOV5m	16	93.3	77.9	63.9	78.4
	30	93.6	81.6	52.0	75.7
YOLOV5l	16	92.6	80.9	65.0	79.5
	30	95.2	81.7	65.1	80.7
YOLOV5x	16	93.8	80.0	59.1	77.6
	30	93.1	82.8	66.3	80.7
YOLOV7	16	96.1	76.8	67.5	80.2
	30	93.9	78.8	67.6	80.1
YOLOV7x	16	95.4	81.0	62.1	79.5
	30	95.7	88.1	60.7	81.5
YOLOV8s	16	92.6	83.2	62.6	79.5
	30	90.9	75.5	66.4	77.6
YOLOV8m	16	92.2	82.8	55.3	77.1
	30	**90.1**	**83.2**	**73.8**	**82.4**
YOLOV8l	16	92.5	76.2	61.8	76.8
	30	93.2	79.2	70.3	80.9
YOLOV8x	16	91.8	77.0	65.8	78.2
	30	93.6	77.3	62.0	77.7

Table:3

We can view predicted images and the related recall curves for the more accurate model between batch 16 and 30 for each architecture by looking at the figures below.

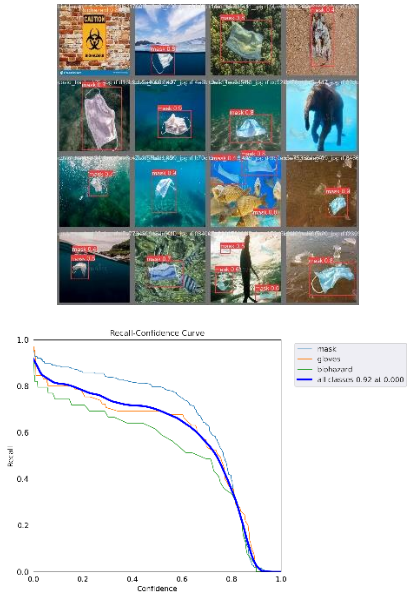

Figure 1: YOLOV5s (B30) Prediction and Recall curve

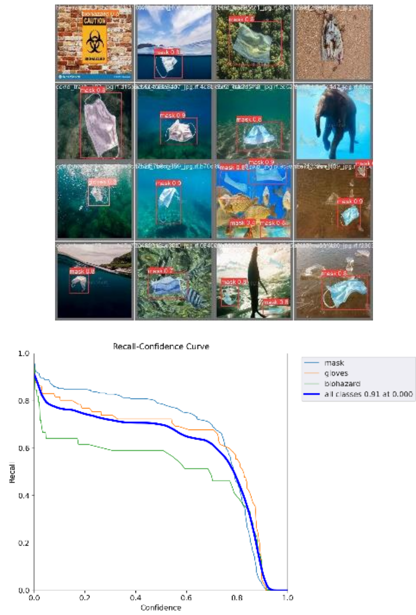

Figure 2: YOLOV5m (B16) Prediction and Recall curve

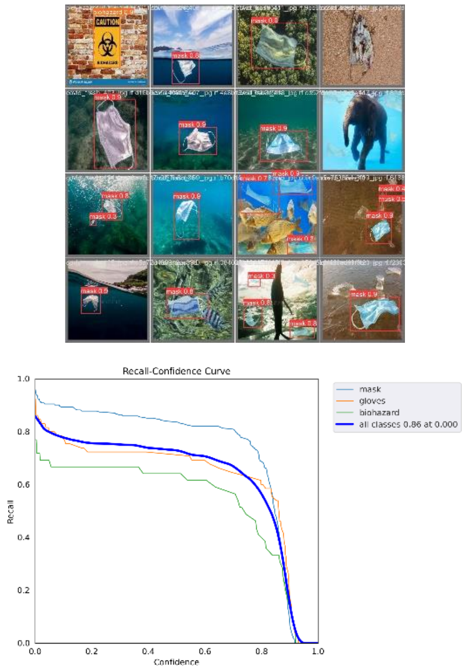

Figure 3: YOLOV5x (B30) Prediction and Recall curve

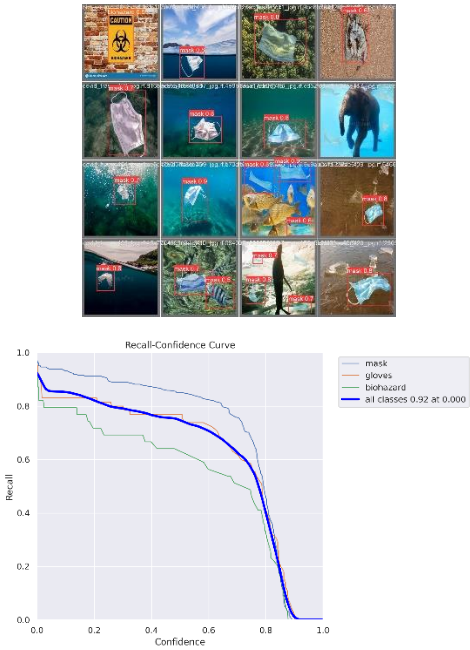

Figure 4: YOLOV5l (B30) Prediction and Recall curve

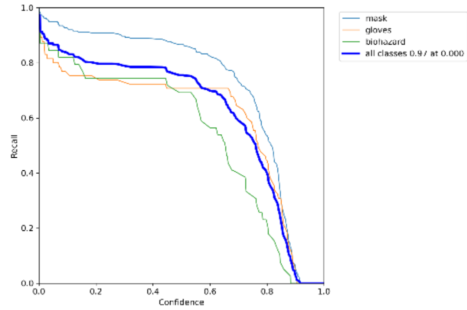

Figure 5: YOLOV7 (B16) Prediction and Recall curve

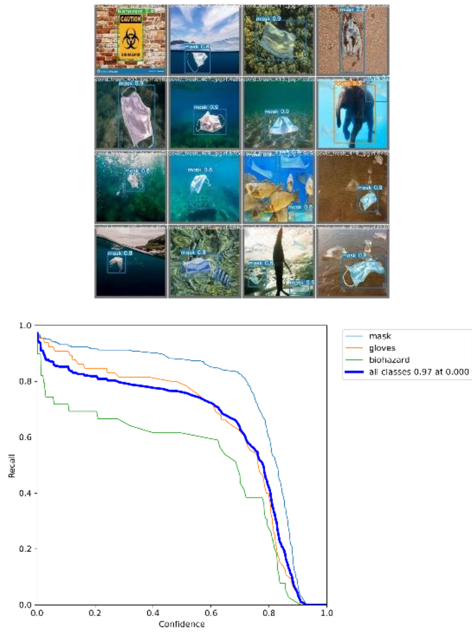

Figure 6: YOLOV7x (B30) Prediction and Recall curve

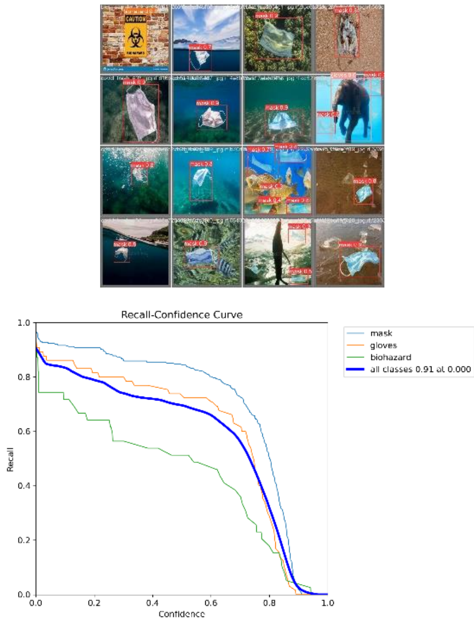

Figure 7: YOLOV8s (B16) Prediction and Recall curve

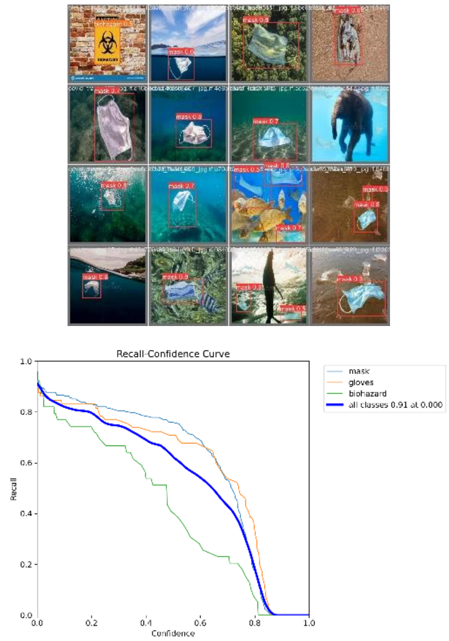

Figure 8: YOLOV8m (B30) Prediction and Recall curve

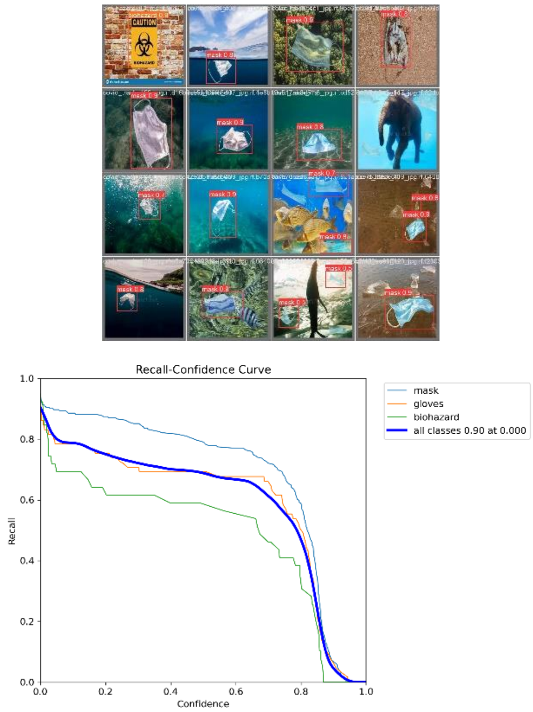

Figure 9: YOLOV8l (B30) Prediction and Recall curve

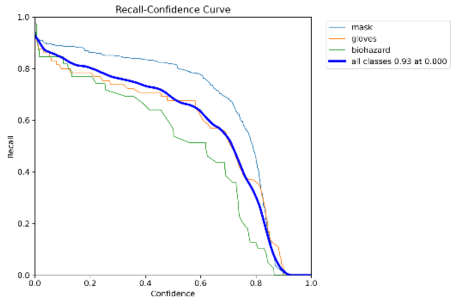

Figure 10: YOLOV8x (B16) Prediction and Recall curve

B. Discussion

In this study, we evaluated and contrasted the effectiveness of a variety of YOLO architectures using the MSG dataset. The findings are presented in the table down below. When assessed with a batch size of 30, the YOLOv5-s architecture performed much better than the other YOLOv5 models, achieving an overall mAP of 81.8%. With an aggregate mAP score of 81.5%, YOLOv7-x performed much better than YOLOv7. When compared to the other YOLOv8 models, the performance of the YOLOv8-m architecture was superior to that of the other YOLOv8 models, with an overall mAP score of 82.4% when the batch size was 30, which was likewise the highest score among all of the other models.

V. Conclusion

We concluded our analysis of the MSG dataset by comparing several YOLO-based architectures. According to the results, mAP was raised by 82.4% using anchor-free YOLOv8-m and a batch size of 30. The results will improve the design of automated systems, which will lessen the likelihood of infection for both patients and medical staff. This study does not go into additional deep learning methodologies or conventional methods for detecting surgical waste, instead concentrating solely on YOLO-based architectures. The performance evaluation is also undertaken using a limited dataset that may not be indicative of the wide variety of surgical waste experienced in practice. To overcome these gaps and learn everything there is to know about finding and disposing of surgical waste, further study is required. Additional datasets might be used in future efforts in this area to improve the precision and reliability of surgical waste identification. Also, as needs for medical waste management tend to change over time, looking into how the chosen YOLO architecture performs in a real-world healthcare context would be useful. Research of this nature might evaluate the system's functionality in the actual world and reveal any potential problems or restrictions.

References

[1] M. Ferdous and S. M. M. Ahsan, "A Computer Vision-based System for Surgical Waste Detection," *International Journal of Advanced Computer Science and Applications,* vol. 13, no. 3, 2022.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.

[3] M. Hammami, D. Friboulet, and R. Kéchichian, "Cycle GAN-based data augmentation for multi-organ detection in CT images via YOLO," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 390-393: IEEE.

[4] P. Adarsh, P. Rathi, and M. Kumar, "YOLO v3-Tiny: Object Detection and Recognition using one stage improved model," in *2020 6th international conference on advanced computing and communication systems (ICACCS)*, 2020, pp. 687-694: IEEE.

[5] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection," *Sustainable Cities and Society,* vol. 65, p. 102600, 2021.

[6] M. Horvat and G. Gledec, "A comparative study of YOLOv5 models performance for image localization and classification," in *Central European Conference on Information and Intelligent Systems*, 2022, pp. 349-356: Faculty of Organization and Informatics Varazdin.

[7] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv preprint arXiv:.02696,* 2022.

[8] J. Chen, J. Mao, C. Thiel, and Y. Wang, "iWaste: Video-based medical waste detection and classification," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 5794-5797: IEEE.

[9] K. Themistocleous, C. Papoutsa, S. Michaelides, and D. Hadjimitsis, "Investigating detection of floating plastic litter from space using sentinel-2 imagery," *Remote Sensing,* vol. 12, no. 16, p. 2648, 2020.

[10] N. M. Kumar *et al.*, "Artificial intelligence-based solution for sorting COVID related medical waste streams and supporting data-driven decisions for smart circular economy practice," *Process Safety and Environmental Protection* vol. 152, pp. 482-494, 2021.

[11] H. Panwar *et al.*, "AquaVision: Automating the detection of waste in water bodies using deep transfer learning," *Case Studies in Chemicaland Environmental Engineering,* vol. 2, p. 100026, 2020.

[12] N. Mehendale, V. Sule, C. Tamhankar, S. Kaveri, and N. Lakade, "Computer vision based medical waste separator," *Available at SSRN 385780,* 2021.

[13] A. Buragohain, B. Mali, S. Saha, and P. K. Singh, "A deep transfer learning based approach to detect COVID‐19 waste," *Internet Technology Letters,* vol. 5, no. 3, p. e327, 2022.

CONTRIBUTION TABLE

	Shakib Sadat Shanto	**Zishan Ahmed**	**Samin**	**Contribution (%)**
	20-43074-1	***20-42085-1***	***20-42408-1***	
Topic Selection	35%	35%	30%	100 %
Analysis	40%	30%	30%	100 %
Related Paper Reading	40%	30%	30%	100 %
Code Implementation	30%	40%	30%	100 %
Paper Writing	35%	35%	30%	100 %

➔ **GitHub Link**