Real-Time Harassment Detection in Indian Workplaces Using Advanced AI and Computer Vision

Nikhileswara Rao Sulake[1],Sai Manikanta Eswar Machar[1], Challapalli Srinivasu[1] , Bhatraju Mahalakshmi Rao[1]

[1] Department of CSE, RGUKT Nuzvid, Andhra Pradesh, India
nikhil01446@gmail.com

Abstract. Workplace harassment affects 31% of Indian working women, as reported by the National Commission for Women [1], yet underreporting persists due to social stigma and lack of evidence. This paper proposes Swetcha, an automated AI framework designed to detect and classify harassment in real time by integrating advanced computer vision with workplace surveillance systems. Leveraging transfer learning, the YOLOv8 [2] architecture is fine-tuned on a custom dataset comprising Indian workplace scenarios, including both normative interactions and harassment incidents (e.g., unwarranted physical proximity, aggressive gestures). The system employs skeletal pose estimation to analyze joint movements and facial expression recognition to identify distress cues, enabling context-aware classification of events as Normal or Harassment with a confidence threshold of ≥85%. Live CCTV footage is processed frame-by-frame, with flagged incidents stored securely as timestamped clips for evidentiary purposes. A critical feature is its integration with organisational POSH (Prevention of Sexual Harassment) committees, where alerts are escalated automatically alongside contextual metadata to expedite resolution under India's POSH Act (2013) [3]. Privacy safeguards, such as dynamic blurring of non-involved individuals, ensure ethical compliance. By merging real-time analysis with proactive governance, Swetcha fosters safer workplaces through timely intervention, accountability, and evidentiary transparency, aligning with India's digital governance and workplace safety mandates.

Keywords: Women Safety, Harassment Detection, Computer Vision, CCTV

1. Introduction

Workplace harassment against women in India persists due to power imbalances, social stigma, and weak enforcement of laws like the POSH Act (2013). Despite 46.58% of women facing harassment [4], over 70% of cases go unreported. With only 18% of organizations ensuring compliance, systemic failures create a culture of impunity. Existing surveillance methods are ineffective, and courts demand timestamped, context-rich evidence, underscoring the need for a robust, context-aware framework.

This paper introduces Swetcha, an AI-powered surveillance system designed to address these gaps through real-time, harassment detection in workplace environments. Leveraging transfer learning on the YOLOv8 architecture, Swetcha analyzes CCTV streams to identify harassment markers , abnormal body postures, and facial distress cues in controlled trials. The system employs a hybrid approach: skeletal pose estimation maps joint movements to flag potential threats, while a library named DeepFace [5] is used, which classifies facial expressions using a dataset curated from diverse open resources. By automating detection and evidence preservation, Swetcha shifts workplaces from reactive damage control to proactive prevention, a critical step toward fulfilling the POSH Act's vision.

2. Methodology

The methodology for Swetcha's development is structured into three stages: Dataset Preparation, Model Training and Fine-Tuning and Real-Time Detection. Each stage is designed to ensure robustness, scalability, and ethical compliance in workplace surveillance.

2.1 Dataset Preparation

The dataset was curated from publicly available online sources, comprising both videos and images. The collected data was manually annotated and categorized into two primary classes: Harassment and Normal. Videos depicting harassment scenarios were trimmed and saved into the "Harassment" folder, while normal scenarios were stored in the "Normal" folder.

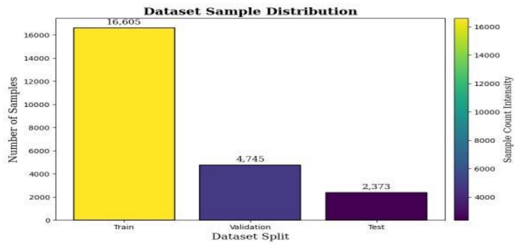

Fig. 1: Dataset split over train, test and valid. This includes training the model with enough resources for it's generalizing capabilities.

2.2 Model Training and Fine-Tuning

The YOLOv8 model was selected as the base architecture due to its state-of-the-art performance in object detection and classification tasks. Transfer learning was employed to fine-tune the pre-trained YOLOv8 model using the custom dataset (see Table 1). The model was trained for 10 epochs, allowing it to learn the distinguishing features between harassment and normal scenarios. During training, the model was optimized to detect and classify individuals as male or female, analyze key joint points, and interpret facial expressions.

Table 1. This table contains the details of the model

Metrics	Value
Total Layer	141
Total Parameters	15.77 million
Computational Overhead	41.9 GLOPS

2.3 Real-Time Harassment Detection

The trained model integrates with a live video processing system that continuously analyzes incoming camera footage. The system processes each frame by examining body joint positions and facial expressions of people in view. Based on these features, it classifies interactions as either concerning or normal behavior. When concerning behavior is detected, the system securely stores the frames with associated timestamps. Each detection includes a confidence score to indicate the model's certainty level, ensuring reliable and transparent monitoring (see Fig 1)

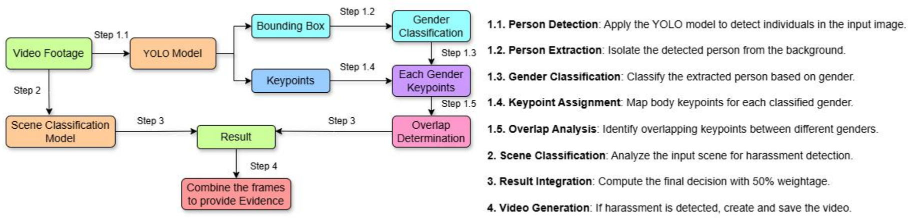

Fig 1. The complete algorithm of the project, from taking the video footage to saving the harassment detected frame into a video for evidence purposes.

3. Experimentation

We initially developed a pipeline integrating pose detection, facial expression recognition, and gender classification to identify harassment in video frames. This approach first classified the gender of individuals in the frame, utilized pose detection to analyze key joint points, and identified abnormal overlaps between individuals. If such overlaps were detected, the system analyzed the facial expressions of females to assess signs of distress or discomfort. Frames were then labeled as either "Harassment" or "Normal" based on these evaluations.

However, this method proved to be overly complex, resource-intensive, and inadequate for addressing all possible scenarios. Consequently, we transitioned to a more efficient and scalable solution using the YOLOv8 model. Leveraging transfer learning, we trained the YOLOv8 model on a custom dataset comprising images and videos sourced from various origins. The dataset was meticulously curated to include diverse harassment scenarios, enabling the model to generalize effectively across different contexts. The trained model was optimized for real-time frame classification as either "Harassment" or "Normal," utilizing the features learned during training.

This approach not only streamlined the pipeline but also enhanced system performance and its ability to address a broader range of harassment cases.

4. Project Pipeline

The proposed pipeline for the harassment detection system is designed to process video footage in real time, classifying each frame. The system begins with the input of live video footage, where frames are continuously extracted for analysis. Each frame is preprocessed to meet the YOLOv8 model's input requirements. The YOLOv8 model, fine-tuned using transfer learning, then performs feature extraction and classification by identifying individuals, analyzing their interactions, and detecting potential harassment scenarios. The model evaluates features such as body positioning, proximity, and contextual cues to accurately classify the frames.

Fig. 2: These 4 image outputs showcase the working of the backend keypoint logic between two different gender

Frames classified as "Harassment" are labeled with a confidence score and stored. The video clip of the flagged frames is generated, complete with timestamps, and forwarded to appropriate authorities, such as POSH cells, to facilitate timely review and intervention. The system also maintains a systematic log of flagged incidents for transparency and potential legal proceedings. By automating the detection and reporting process, the pipeline streamlines real-time monitoring, reducing manual surveillance efforts while ensuring a proactive approach to addressing harassment cases in workplace environments.

5. Results

The YOLOv8 model was trained for a binary classification task using a dataset as mentioned. Training was conducted over 10 epochs with a resolution of 128x128 pixels, and the model demonstrated strong performance in accurately classifying the dataset. Key metrics include a Top-1 accuracy of 99.8% and a Top-5 accuracy of 100%, with the final training loss recorded as 0.2084. The model architecture, consisting of 141 layers and 15.77 million parameters, efficiently managed the task with minimal computational overhead (41.9 GFLOPs). Transfer learning from pre-trained weights facilitated rapid convergence, particularly noticeable in the first epoch, while the AdamW optimizer with a learning rate of 0.000714 effectively minimized loss. Additionally, Automatic Mixed Precision enhanced training speed without compromising accuracy (see Table 2).

Table 2. This table contains the summary of the results

Metrics	Value
Training Images	16.605
Validation Images	4745
Test Images	2373
Image Resolution	128
Top-1 Accuracy	99.8
Top-5 Accuracy	99.9
Training Loss	0.2084

The model's performance during validation confirmed its ability to learn discriminative features to differentiate between the two classes, even with low-resolution images. These results underscore the robustness and scalability of YOLOv8 for classification tasks in resource-constrained environments. The successful outcomes demonstrate its potential for real-world applications where computational efficiency and accuracy are critical. The working of the keypoints logic is shown in Fig 1, the final output of the pipeline is shown in Fig 3 and a real case test in Fig 4.

Fig. 3: This is the main output of the complete pipeline, where the two different gets segmented and based on the keypoints and female face emotion, it is classified as harassment or not.

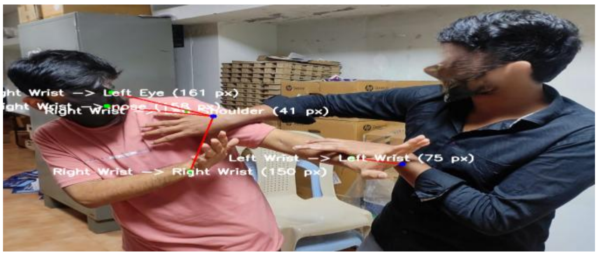

Fig. 4: A Real life test case where one male is set as female for testing purposes (Faces are blured)

6. Conclusion

Workplace harassment remains a critical issue that demands innovative and effective solutions. Our proposed system leverages the power of AI and computer vision to provide real-time monitoring and detection of harassment scenarios, creating a safer and more inclusive work environment. By employing the YOLOv8 model with transfer learning, we have developed an efficient and scalable solution capable of classifying video frames with high accuracy. The system not only detects and flags potential harassment incidents but also ensures timely reporting to the appropriate authorities, enabling quick action and legal compliance. With its ability to continuously analyze workplace environments and provide actionable insights, this solution has the potential to significantly reduce harassment cases and promote a culture of accountability and safety.

7. Future Scope

Future enhancements aim to boost the system's robustness and adaptability across diverse environments by integrating advanced emotion recognitionand expanding audio analysis to identify verbal harassment, enhancing situational awareness. The system will be optimized for crowded public spaces (e.g., parks, malls) through multi-camera inputs for comprehensive tracking and unsupervised learning to detect anomalies. Transitioning to a cloud-based platform will improve scalability and remote monitoring, while proactive measures like real-time alerts could deter incidents before escalation, fostering safer, more dignified spaces.

8. Data Availability Statement

The dataset, consisting of approximately 24,000 images from online sources and real-life dummy samples, is restricted for research use. Access may be granted upon request by contacting the corresponding author

References

1. National Commission for Women. Annual Report 2022–2023. NCW, 2023, https://ncwapps.nic.in/pdfReports/AnnualReport2022_2023_Eng.pdf
2. Jocher, Glenn, et al. "YOLOv8: State-of-the-Art Object Detection Model." Ultralytics, 2023, https://docs.ultralytics.com/models/yolov8/
3. Government of India. The Sexual Harassment of Women at Workplace (Prevention, Prohibition and Redressal) Act, 2013. Ministry of Women and Child Development, 2013, https://www.indiacode.nic.in/bitstream/123456789/2104/1/A2013-14.pdf
4. Singh MM, Patnaik SK, et al. PK. Sexual Harassment: A Growing Concern for Women in Indian Healthcare Industry. Int J Res Foundation Hosp Healthc Adm 2014;2(1):41-49.
5. Serengil, S., & Özpınar, A. (2024). A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules. Bilişim Teknolojileri Dergisi, 17(2), 95-107. https://doi.org/10.17671/gazibtd.1399077