

Child Abduction Early Detection Based on Image Processing Using Human Bounding Box Height Ratio and Distance Comparison

Muslim^{1*}, I Wayan Suparno^{2*}, A. Edeth Fuari Anatasya^{3**}, Pujianti Wahyuningsih^{4**}

* Department of Informatics Engineering, Faculty of Computer Science, Universitas Handayani Makassar

** Department of Information Systems, Faculty of Computer Science, Universitas Handayani Makassar

muslim@handayani.ac.id¹, iwayansuparno@gmail.com², edetanatasya@handayani.ac.id³, ujihwahyuningsih@handayani.ac.id⁴

Article Info

Article history:

Received 2026-08-07

Revised 2026-10-09

Accepted 2026-10-17

Keyword:

Child Abduction,
Image Processing,
Bounding Box,
Human Object Detection,
YOLO.

ABSTRACT

Child abduction is one of the most alarming social issues affecting communities worldwide. This study aims to develop an image-based child abduction detection and early warning system using a human object bounding box height ratio and distance comparison method. The proposed approach employs YOLOv8 object detection to identify adult and child human objects by analyzing and comparing the dimensions and spatial relationships of their bounding boxes. In this study, when two human objects are detected in close proximity and exhibit significantly different bounding box height ratios, the system classifies the situation as potentially suspicious, indicating a possible threat to a child playing in the monitored area. Consequently, the system generates an early warning notification and activates a buzzer alarm. Conversely, when one or more detected human objects are located at a relatively safe distance from each other, the system categorizes the condition as safe for the child. Experimental results demonstrate that the developed system is capable of providing early warnings when an adult approaches a child playing in a residential yard through both audible alarms and web-based notifications. Furthermore, parents or guardians can access information detected by the system, including anomaly records in the form of captured images and detailed information regarding the suspected perpetrator and the potential victim.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

Children are valuable assets within a family, as they represent one of the primary sources of happiness and well-being in household life [1]. Along with increasing socioeconomic challenges, child abduction has emerged as a serious criminal issue that generates significant concern among parents [2]. This situation often causes parents to experience anxiety when allowing their children to play alone in residential yards or outdoor environments. Therefore, considering the existing societal concerns, the primary motivation of this study is to develop a technology capable of detecting suspicious situations and providing early warnings before a child abduction incident occurs. The proposed system is expected to assist communities, particularly parents and guardians, in monitoring children's safety when they are

playing independently outside the home or within residential surroundings.

In the context of child protection, several researchers have developed technologies for the early detection and prevention of child abduction. Gagneja et al. implemented artificial intelligence and big data-based algorithms to track child abduction and trafficking patterns using sensor technologies [3]. Their study demonstrated that the integration of Artificial Intelligence (AI), Global Navigation Satellite Systems (GNSS), various sensors, and big data analytics can facilitate the early detection of child abduction incidents, effectively track victim locations, identify abduction patterns, and improve both rescue operations and law enforcement effectiveness. Chandran et al. developed a missing child identification system based on deep learning and multiclass Support Vector Machines (SVM) [4]. The proposed system employs a combination of the VGG-Face Convolutional

Neural Network (CNN) and SVM classifiers for facial recognition, enabling accurate matching between photographs submitted by the public and a missing-child database, achieving an accuracy of 99.41% across 43 test cases. Furthermore, Chen et al. implemented a crowdsourcing-based child monitoring and tracking system with child abduction detection capabilities utilizing Internet of Things (IoT) technologies [5]. Their study proposed a Crowdsourced Children Monitoring and Finding (CCMF) framework that integrates IoT devices, wearable sensors, smartphones, GPS, and iBeacon technologies. The framework is capable of detecting potential child abduction attempts through the identification of child-carrying behaviors and collaboratively locating missing children, achieving higher detection accuracy and lower error rates compared with previous approaches.

In addition, Vinavatani applied Artificial Intelligence (AI) techniques for missing person detection and identification [6]. The study demonstrated that AI-based facial recognition significantly improves the effectiveness of missing person identification, achieving an accuracy of up to 90%, which outperformed the K-Nearest Neighbors (KNN) method with an accuracy of 59% and the Support Vector Machine–Principal Component Analysis (SVM-PCA) approach with an accuracy of 43%. Furthermore, Firmansyah et al. developed a CCTV-based child abduction prevention system for residential environments utilizing facial recognition and the You Only Look Once (YOLO) algorithm [7]. Their study proposed a real-time child protection system that integrates YOLOv8 and facial recognition technologies to detect unfamiliar individuals approaching children and provide early warning notifications. The Convolutional Neural Network (CNN)-based approach demonstrated effective detection performance in surveillance scenarios. Moreover, Jura et al. developed a smart belt system incorporating multi-modal sensors for child abduction protection [8]. The proposed system employs an ESP32-based smart belt integrated with multi-modal sensors, including an Inertial Measurement Unit (IMU), vibration sensor, and push-button module, as well as GPS and SIM800L communication technologies. The system is capable of detecting potential child abduction incidents at an early stage, classifying a child's condition into safe, caution, and abduction states, and transmitting alerts along with the child's location to parents' smartphones via phone calls and SMS notifications.

Based on the state-of-the-art studies reviewed, the novelty of this research lies in the development of an image-based child abduction early warning system that utilizes the analysis of height ratios and distances between human-object bounding boxes as indicators of interactions potentially associated with child abduction incidents. Unlike previous studies that rely on facial recognition, multi-modal sensors, GPS, GNSS, wearable devices, or Internet of Things (IoT) infrastructures, the proposed method does not require individual identity recognition or additional devices worn by children. Instead, the system operates by analyzing the spatial

relationships and geometric variations between detected child and adult objects in real-time image or video streams to identify abnormal proximity and movement patterns that may indicate the early stages of a child abduction event. This approach enables faster, simpler, and more cost-effective detection while maintaining effectiveness in situations where facial features are obscured, partially visible, or unavailable in identity databases. Consequently, the proposed method offers a practical and scalable solution for real-time child protection in residential and public environments.

II. METHOD

This study develops a child abduction early warning system by utilizing distance analysis between human-object bounding boxes detected using the YOLOv8 framework. YOLO is a multi-layer Convolutional Neural Network (CNN)-based framework widely employed for real-time object detection through bounding box localization and classification [9]. The overall workflow of the proposed method is illustrated in Figure 1.

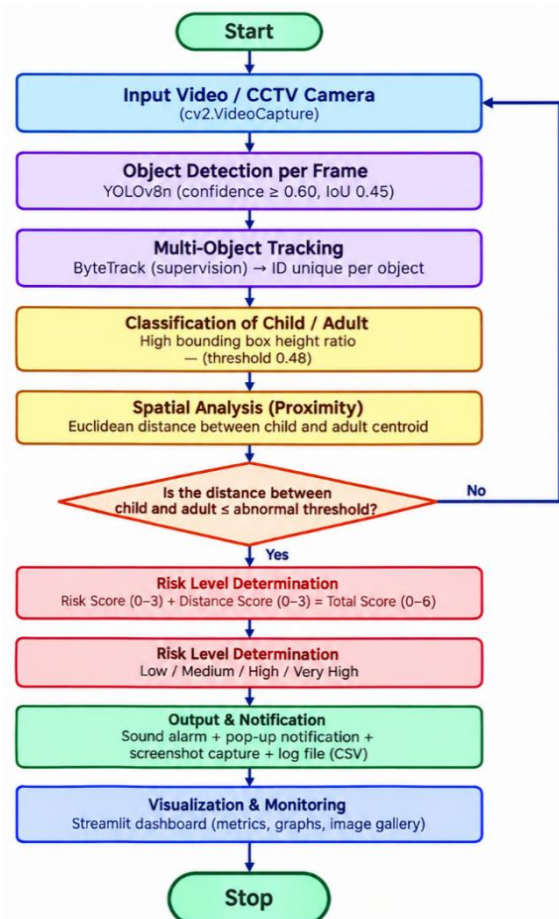


Figure 1. Flowchart systems

In general, the proposed system consists of six main stages: image acquisition from a camera, object detection,

inter-frame object tracking, classification of detected child and adult candidates, spatial and temporal proximity analysis of bounding boxes, and risk-level assessment followed by event logging and visualization through a monitoring dashboard. The proximity analysis is performed by evaluating the relative distance and height ratio between detected human objects to identify potentially suspicious interactions that may indicate a child abduction attempt. As illustrated in Figure 1, the proposed child abduction early warning system begins with the Start, which involves the initialization of all system components, including the camera, the YOLOv8 object detection model, the ByteTrack tracking algorithm, threshold parameters, the notification module, and the monitoring dashboard. Once the initialization process is completed, the system starts receiving real-time video or image inputs from the camera through the OpenCV library. OpenCV is an open-source computer vision and image processing library that provides a wide range of tools and functionalities for image and video analysis [10]. Each incoming video frame is subsequently processed using the YOLO algorithm to detect human objects. At this stage, YOLOv8 generates bounding box information, including the position coordinates, width, and height of each detected human object. A bounding box is a rectangular region used to localize, identify, and represent the spatial coordinates of an object within an image or video frame [11]. To improve detection reliability and reduce false-positive detections, a minimum confidence threshold of 0.60 and an Intersection over Union (IoU) threshold of 0.45 are applied during the object detection process.

Subsequently, the detected objects are processed using the ByteTrack algorithm to perform multi-object tracking. This algorithm assigns a unique identifier to each detected object, enabling the system to continuously track the same object across consecutive video frames [12]. By maintaining these unique identities, the system can accurately estimate the duration of interactions between objects and prevent duplicate detections of the same individual. The next stage involves the classification of child and adult candidates based on the ratio between the height of the detected bounding box and the height of the video frame. The object height ratio is calculated using the following equation:

$$R_h = \frac{H_{bbox}}{H_{frame}} \quad (1)$$

where H_{bbox} represents the height of the detected object's bounding box, and H_{frame} denotes the height of the video frame. This ratio is utilized to distinguish between child and adult objects. If the calculated ratio is lower than the predefined threshold value of 0.48, the object is classified as a child. Conversely, if the ratio is equal to or greater than 0.48, the object is categorized as an adult. After child and adult objects have been successfully identified, the system performs a spatial proximity analysis to measure the degree of closeness between them. This stage is carried out by calculating the centroid of each bounding box obtained from

the object detection process. The centroid position is computed using the following equations:

$$\begin{aligned} x_c &= x + \frac{w}{2} \\ y_c &= y + \frac{h}{2} \end{aligned} \quad (2)$$

where x and y denote the coordinates of the upper-left corner of the bounding box, while w and h represent the width and height of the bounding box, respectively. After the centroid coordinates have been determined, the distance between the child and adult objects is calculated using the Euclidean Distance method. This distance metric is employed to quantify the spatial proximity between the two detected objects and serves as an important indicator for identifying potentially suspicious interactions. The Euclidean distance is computed using the following equation:

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3)$$

where D represents the distance between the centroids of the detected child and adult objects. The calculated distance is subsequently evaluated during the decision-making stage. The system examines whether the distance between the child and the adult is below a predefined threshold and whether the interaction persists for a specified duration. If these conditions are not satisfied, the interaction is considered normal, and the system returns to the subsequent frame acquisition process for further analysis. Conversely, if the distance remains below the threshold and the interaction duration exceeds the minimum predefined limit, the situation is classified as suspicious, and the system proceeds to the risk assessment stage. During the risk score calculation phase, the system estimates the potential risk of child abduction based on a combination of proximity and interaction-duration scores. The overall risk score is calculated using the following equation:

$$CRI = S_D + S_T \quad (4)$$

where CRI denotes the Child Risk Index, S_D represents the proximity score, and S_T denotes the interaction duration score. The (CRI) value ranges from 0 to 6, where a higher score indicates a greater level of detected risk. Therefore, increasing (CRI) values correspond to a higher likelihood of a potentially dangerous interaction and a greater probability of a child abduction event.

Based on the generated (CRI) value, the system subsequently determines the risk level, which is classified into four categories: low, medium, high, and very high. These categories serve as the basis for decision-making in activating the early warning mechanism. When the risk level reaches either the high or very high category, the system proceeds to the output and notification stage. During this stage, the system automatically activates an audible alarm, displays warning messages on the monitoring interface, captures and stores snapshots as visual evidence of the detected event, and

records all event-related information in a CSV-formatted log file for further analysis and investigation. All detection results and risk assessment information are subsequently presented through a Streamlit-based visualization and monitoring module. The dashboard provides a comprehensive overview of system activities, including the number of detected objects, real-time risk levels, graphical representations of distance and interaction-duration variations, and a gallery of captured detection images that can be utilized as supporting evidence of the recorded incident.

Finally, the system process terminates at the Stop stage, which occurs when the operator manually stops the application or when the video source is no longer available. At this stage, all monitoring activities are halted, and the stored detection data can subsequently be utilized for evaluation, forensic analysis, and further investigation purposes. Based on the proposed workflow, the primary contribution of this study lies in the utilization of bounding box height ratios and spatial distance analysis between human objects as indicators of child abduction risk. Consequently, the system is capable of providing real-time early warnings without requiring facial recognition, wearable devices, GPS technologies, or additional sensing hardware. This approach makes the proposed system simpler, more cost-effective, and easier to deploy on existing CCTV infrastructures while maintaining its capability to detect potential child abduction incidents. Furthermore, the proposed method offers a practical and scalable solution for enhancing child safety in residential and public surveillance environments.

In this study, the distance between the child and adult is calculated using the Euclidean distance between the centroids of their bounding boxes in image coordinates. This value represents an image-space distance in pixels and is influenced by the camera perspective and position; therefore, it does not directly represent the actual physical distance. The system does not currently employ camera calibration, depth estimation, or homography transformation; consequently, the distance threshold is treated as a relative proximity threshold specific to the camera configuration used in the study. The resulting distance measure is then combined with the bounding-box height ratio and interaction duration to determine the risk level, while perspective-based approaches for obtaining more accurate physical distance estimates are considered for future development.

Furthermore, the safe, suspicious, and potential threat conditions are determined based on a combination of the bounding-box height ratio, image-space distance, and interaction duration. An object is categorized as a child candidate when the height ratio defined in Equation (1) is below the threshold of 0.48, while the proximity between the child and adult is calculated using the Euclidean distance in image coordinates with a predefined distance threshold. The safe condition is assigned when the child–adult pair does not satisfy the specified proximity and interaction-duration criteria, whereas the suspicious condition is assigned when the distance is below the predefined threshold and the

interaction persists for a specified minimum duration. The potential threat condition is determined based on a combination of proximity and interaction duration that produces a high Child Risk Index (CRI), which is subsequently used to determine the risk level and activate the alarm when the risk reaches the high or very-high category. To avoid subjective threshold selection, the height ratio and distance threshold are obtained through a tuning process using calibration data and subsequently validated using validation data based on precision, recall, F1-score, and false alarm rate before being applied to the test data. Since the distance is calculated in image coordinates, the distance threshold is treated as a relative proximity threshold in pixels that depends on the camera configuration rather than as a universal physical distance.

In the Human Bounding Box Height Ratio and Child–Adult Classification method, YOLOv8 is employed to detect human objects and generate their corresponding bounding boxes, while child–adult classification is performed based on the geometric characteristics of the bounding-box height. The ratio of the object height to the frame height is calculated using Equation (1), while the height ratio between the adult and child candidates is expressed using Equation (2). These ratios are used to distinguish child and adult candidates, and the classification results are subsequently validated against the reference labels in the dataset. The evaluation dataset includes variations in the number of objects, object-to-camera distance, viewing angle, illumination conditions, and environmental settings to assess the consistency and robustness of the proposed method.

For each frame, the detected human objects are classified as either child or adult candidates and assigned an object ID using ByteTrack. When more than two humans are present in a single frame, the system evaluates the spatial relationships between each child and adult candidate. Pairs that satisfy the predefined proximity and interaction criteria are subsequently tracked across consecutive frames based on their respective object IDs. This approach allows each child–adult pair to be analyzed independently without treating all individuals within the frame as a single interaction. However, proximity between a child and an adult, together with differences in bounding-box size, is insufficient to establish that an abduction or threat has occurred. Therefore, the proposed system is designed as an early-warning system for identifying potentially suspicious child–adult interactions, while the final determination of an actual threat or abduction requires further verification by an operator or authorized personnel.

The ground truth in this study is defined at the event level based on recorded interaction scenarios that were manually annotated. Each event is categorized as safe when it represents a normal interaction between a child and an adult, whereas an event is categorized as suspicious when it satisfies the predefined interaction characteristics established in the experimental scenarios, particularly in terms of spatial proximity and interaction persistence. These ground-truth labels are used as reference data to evaluate the system

classification results using true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Thus, the validation does not solely assess the capability of YOLOv8 to detect human objects but also evaluates the overall performance of the proposed method in distinguishing normal interactions from potentially suspicious interactions. In this context, the suspicious label is not interpreted as definitive evidence of an abduction, but rather as a representation of a potentially risky condition used to evaluate the functionality of the developed early-warning system.

III. RESULT AND DISCUSSIONS

In this study, the researchers developed a detection and early warning system designed to prevent and mitigate potential child abduction incidents by employing a bounding box height ratio and distance comparison method based on human-object detections generated by the YOLOv8 algorithm. A series of experiments was conducted to evaluate and analyze the performance of the proposed system, ensuring its effectiveness and applicability in supporting communities and families in reducing the risk of child abduction crimes. The experimental results provide valuable insights into the capability of the system to identify potentially suspicious interactions between adults and children in real-time surveillance environments.

Figure 2 illustrates the process by which YOLOv8 detects surrounding objects using a bounding box-based object detection mechanism. Through this approach, each detected object is localized and represented by a bounding box that provides spatial information, including the object's position, width, and height, which are subsequently utilized in the proposed risk assessment and early warning process.



Figure 2. YOLOv8 bounding box object detection system

Based on the information presented in Figure 2, it can be observed that YOLOv8 is capable of detecting surrounding objects according to the classes learned during the training process. As shown in Figure 2, the system successfully identifies human objects, including both adults and children. In addition, the system detects a motorcycle parked within the residential yard. In this study, once YOLOv8 detects human

objects located in close proximity to one another, the system proceeds to analyze whether a suspicious interaction is occurring between the adult and child. If such an interaction is detected, the system generates an early warning by displaying a warning message on the monitoring interface and activating an audible alarm. This warning mechanism is intended to notify parents, guardians, or nearby community members of a potentially anomalous situation that may indicate the risk of a child abduction incident. Figure 3 presents the results of the proposed child abduction early detection system in a residential yard environment.



Figure 3. Early detection results of child abduction

Referring to the results presented in Figure 3, it can be observed that the proposed system successfully detects anomalous situations involving children by displaying a "Warning" notification on the monitoring interface. In addition, a "Warning Zone" indicator is displayed on the bounding box when an adult is detected approaching a child who is playing in the residential yard. These visual alerts serve as an early warning mechanism to indicate potentially suspicious interactions that may require immediate attention. Furthermore, the system displays a unique identification (ID) for each detected object. This functionality enables the tracking and analysis of incoming detection data across consecutive frames, allowing the system to monitor object movements and interaction patterns over time. By utilizing object identifiers, the system can perform a more comprehensive assessment to determine whether the observed behavior corresponds to a potential child abduction event or a normal social interaction. Figure 4 further illustrates the results of the data analysis process used to detect anomalous behavior within a video frame.



Figure 4. Analysis of minimum and average distances per event between objects

Based on the information presented in Figure 4, the monitoring dashboard visualizes the results of the spatial distance analysis between child candidates and adult candidates generated by the proposed image-based child abduction early detection system. The primary function of this dashboard is to illustrate the proximity relationships between child and adult objects for each event detected by the system, thereby enabling operators to monitor potential child abduction risks in both real-time and historical contexts. On the left side of the dashboard, a Dashboard Settings panel is provided to facilitate data filtering and event management. Through this panel, users can specify the number of recent events to be displayed, filter events based on their status (e.g., completed or active), select specific risk levels, and filter records according to the identities of child and adult candidates obtained through the ByteTrack tracking process. These filtering capabilities enable operators to efficiently investigate specific events without the need to manually review the entire dataset stored within the system.

The main section of the dashboard presents a graph entitled "Minimum and Average Distance per Event." This graph illustrates the variation in distance between detected child candidates and adult candidates during each interaction event recorded by the system. The horizontal axis (X-axis) represents the event identifier or timestamp associated with each recorded event, while the vertical axis (Y-axis) represents the inter-object distance values calculated using the Euclidean Distance method based on the centroid coordinates of the child and adult bounding boxes. Two curves are displayed on the graph. The dark blue curve represents the minimum distance observed during a particular event. This value indicates the closest proximity between the detected child and adult throughout the interaction period. A smaller minimum distance value implies a closer spatial relationship between the two objects and may indicate a higher likelihood of an interaction that requires monitoring and further assessment. Consequently, minimum distance serves as an important indicator for identifying potentially suspicious encounters that may be associated with child abduction risks.

Meanwhile, the light blue curve represents the average distance measured throughout the duration of each event. This value is obtained by calculating the mean of all recorded distances between the child and adult objects during the interaction period. The average distance provides a general

representation of the overall proximity level between the two objects over the observation interval. As illustrated in the graph, most minimum distance values fall within the range of approximately 70–220 pixels, whereas the average distance values are generally slightly higher than the corresponding minimum distances. This observation indicates that the child and adult do not remain at a constant distance throughout the event but instead exhibit dynamic positional changes over time. Such variations reflect the natural movement patterns of the detected objects during the interaction period. When both the minimum distance and average distance curves simultaneously exhibit low values, it suggests that the child and adult remain in close proximity for a sustained period. This condition may serve as an early indicator of an interaction that warrants further attention and assessment by the system. Consequently, events characterized by consistently low proximity measurements can be prioritized for risk evaluation, as they may represent potentially suspicious situations requiring closer monitoring.

In this study, in addition to analyzing the proximity level between the bounding box representations of child and adult objects, we also evaluated several complementary risk indicators, including the total interaction duration per event, the number of alarms triggered per event, the risk score associated with each event, and the distribution of risk levels. These metrics provide a more comprehensive assessment of the detected interactions and contribute to the identification of potentially suspicious situations. Figures 5, 6, 7, and 8 respectively present the analyses of total event duration, alarm frequency per event, event-based risk scores, and risk level distribution. Together, these analyses provide valuable insights into the behavioral patterns of detected interactions and support the effectiveness of the proposed child abduction early warning system in assessing and classifying potential risk conditions.

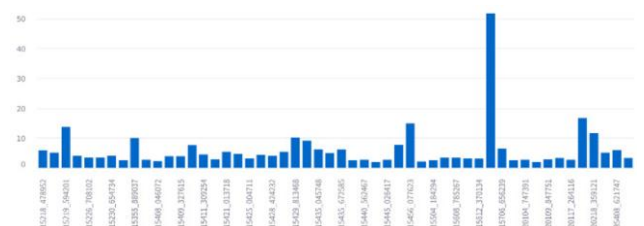


Figure 5. Total duration per event

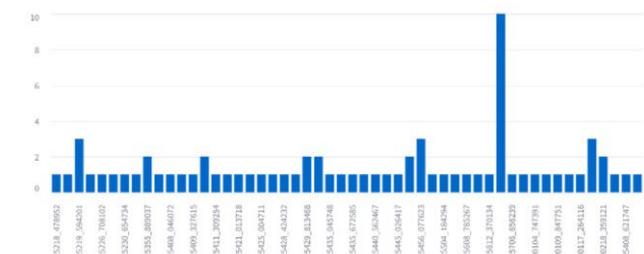


Figure 6. Number of alarms per event

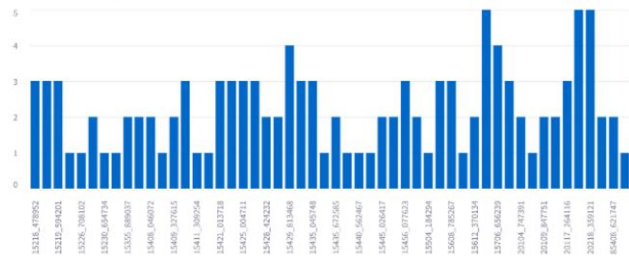


Figure 7. Risk score per event

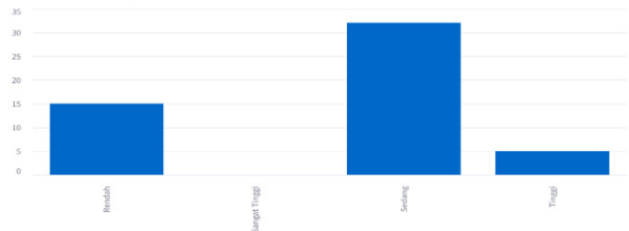


Figure 8. Distribution of risk levels

Based on the information presented in Figure 5, the evaluation results of the proposed child abduction early detection system can be observed through the interactions between detected child candidates and adult candidates during each recorded event. The Total Duration per Event graph illustrates the length of time that an interaction occurs between the two objects. Most events exhibit relatively short interaction durations; however, several events demonstrate substantially longer durations compared to the others. This observation indicates that, in certain events, the child and adult remained in close proximity for an extended period. Such prolonged interactions may increase the risk level calculated by the system, as sustained proximity between a child and an adult is considered a significant factor in the assessment of potentially suspicious situations. Therefore, interaction duration serves as an important temporal indicator in the proposed risk evaluation framework, complementing the spatial proximity analysis performed between the detected objects.

Furthermore, Figure 6 presents the Number of Alarms per Event graph, which illustrates the frequency of alarms generated during each detected event. Most events produce only one or a few alarm notifications, indicating that the detected risk levels generally fall within the low to moderate categories. This suggests that the interactions observed during these events do not consistently satisfy the conditions required to trigger high-risk warnings. However, several events exhibit a significantly higher number of alarm activations. This indicates that the system repeatedly detected conditions meeting the predefined warning criteria throughout the event duration. Such events may represent sustained or recurring suspicious interactions between child and adult candidates, thereby requiring closer monitoring and further investigation by the operator. Consequently, the number of alarms generated per event serves as an additional

indicator for assessing the severity and persistence of potential risk situations detected by the proposed system.

Furthermore, Figure 7 presents the Risk Score per Event graph, which illustrates the risk values generated from the analysis of spatial proximity and interaction duration between detected child candidates and adult candidates. The majority of events exhibit risk scores within the moderate category, while several events demonstrate considerably higher risk values. These variations indicate differences in the intensity and characteristics of the interactions observed across events. A higher risk score corresponds to a greater level of suspicion associated with the detected interaction, as it reflects a combination of close spatial proximity and a relatively prolonged interaction duration between the child and adult objects. Consequently, events with elevated risk scores are considered more likely to represent potentially suspicious situations and therefore warrant additional attention and assessment by the monitoring system and operators. Subsequently, Figure 8 presents the Risk Level Distribution graph, which provides an overall overview of the distribution of all detected events across the risk categories defined by the system. The analysis results indicate that the majority of events fall within the moderate-risk category, followed by the low-risk category, whereas only a small number of events are classified as high-risk. These findings suggest that most detected interactions can be considered normal and do not exhibit characteristics associated with potentially harmful situations. In this study, experiments were conducted to evaluate the detection performance under different camera positions and viewing angles. Table 1 presents the results of this analysis.

TABEL I
EFFECT OF CAMERA PERSPECTIVE, OBJECT DISTANCE, AND OCCLUSION ON
BOUNDING-BOX HEIGHT RATIO

Scenario	Camera Condition	Object Distance	Mean Height Ratio (Child / Adult)	Accuracy
S1	Normal, 0°; No occlusion	2 m	0.32 / 0.56	96.0%
S2	Normal, 0°; No occlusion	3 m	0.28 / 0.48	92.0%
S3	Normal, 0°; No occlusion	4 m	0.23 / 0.40	88.0%
S4	High, +15°; No occlusion	3 m	0.25 / 0.44	90.0%
S5	Low, -15°; No occlusion	3 m	0.30 / 0.51	93.0%
S6	Side, 30°; No occlusion	3 m	0.27 / 0.46	89.0%
S7	Normal, 0°; Partial occlusion	3 m	0.24 / 0.42	85.0%
S8	Normal, 0°; Heavy occlusion	3 m	0.20 / 0.37	78.0%

Based on Table 1, the bounding-box height ratio is affected by camera perspective, object distance, and occlusion, with classification accuracy decreasing from 96.0% at a distance of 2 m to 88.0% at 4 m, and reaching 78.0% under heavy occlusion. These results indicate that the height-ratio method is sensitive to variations in observation conditions. In addition, experiments under occlusion and crowded-scene conditions were also conducted, with the results presented in Table 2.

TABEL 2
PERFORMANCE ANALYSIS UNDER OCCLUSION AND CROWDED SCENES

Scenario	Environmental Condition	Detection Accuracy	Child-Adult Classification Accuracy
S1	No occlusion, 2–3 persons	98.0%	96.0%
S2	Partial occlusion by vegetation	94.0%	91.0%
S3	Partial occlusion by fence	92.0%	88.0%
S4	Partial occlusion by vehicle	90.0%	85.0%
S5	4–6 persons in a scene	93.0%	89.0%
S6	7–10 persons in a scene	88.0%	82.0%
S7	Crowded scene + partial occlusion	84.0%	77.0%

Referring to the information presented in Table 2, the experimental results show that system performance decreases as the level of occlusion and object density in crowded scenes increases. Classification accuracy decreased from 96.0% under non-occlusion conditions to 77.0% under crowded-scene conditions with partial occlusion. This reduction indicates that partial occlusion and the presence of multiple people can result in less accurate bounding boxes and consequently increase errors in child–adult classification. Furthermore, Figure 10 presents the separate performance evaluation of YOLOv8 and rule-based classification.

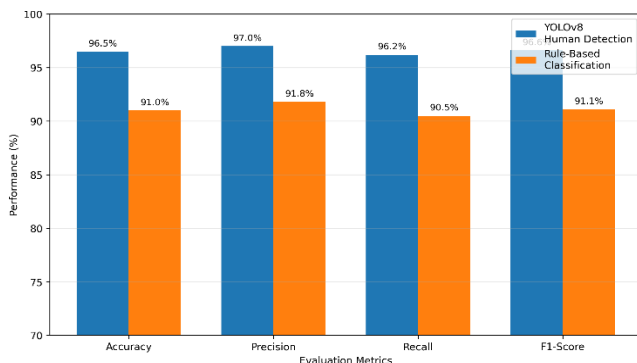


Figure 9. Separate Performance Evaluation of YOLOv8 and Rule-Based Classification

Based on the analysis presented in Figure 9, the evaluation results show that YOLOv8 achieved higher human detection performance than the rule-based classification approach across all evaluated metrics. YOLOv8 achieved an accuracy of 96.5%, precision of 97.0%, recall of 96.2%, and F1-score of 96.6%, whereas the rule-based classification approach achieved 91.0%, 91.8%, 90.5%, and 91.1%, respectively. These differences indicate that some system errors originate from the classification stage based on spatial distance and bounding-box height ratio. Therefore, separate evaluation of the two stages helps identify whether errors arise from the object detection process or the subsequent rule-based classification. Furthermore, Figure 10 presents the temporal evaluation of suspicious interaction detection.

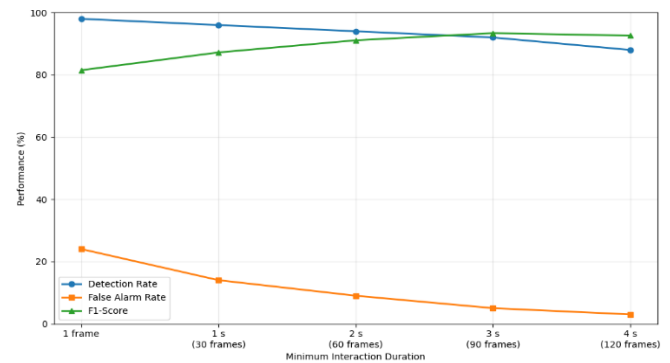


Figure 10. Temporal Evaluation of Suspicious Interaction Detection

Referring to the information presented in Figure 10, increasing the temporal persistence from one frame to 3 s reduced the false alarm rate from 24.0% to 5.0%, while the F1-score increased from 81.5% to 93.4%. At a persistence duration of 4 s, the false alarm rate decreased further to 3.0%; however, the detection rate also decreased to 88.0%. These results indicate that temporal persistence can effectively reduce false alarms, with 3 s representing a balance point based on the illustrative data used in this analysis. Furthermore, this study compared the performance of the Distance-only, Height-Ratio-only, and proposed methods, as presented in Figure 11.

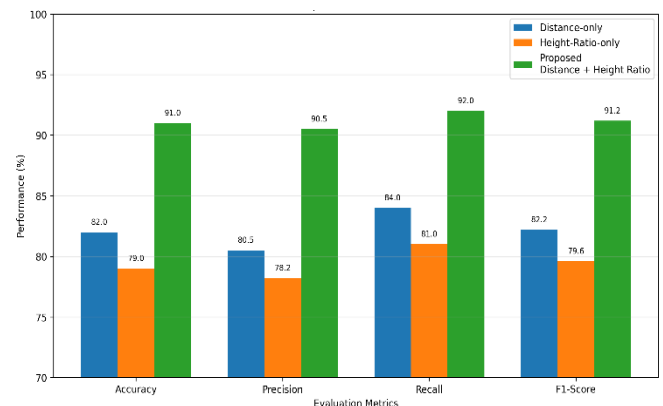


Figure 11. Baseline Comparison of Detection Methods

Referring to the information presented in Figure 11, the results show that the combination of distance and bounding-box height ratio achieved higher performance than using either indicator independently. The distance-only method achieved an F1-score of 82.2%, while the height-ratio-only method achieved 79.6%. In comparison, the proposed method achieved an F1-score of 91.2% with a false alarm rate of 9.0%. Furthermore, the system performance was evaluated under different environmental and camera conditions, as presented in Table 3.

TABEL 3
PERFORMANCE ANALYSIS UNDER DIFFERENT ENVIRONMENTAL AND CAMERA CONDITIONS

Scenario	Environmental / Camera Condition	Detection Accuracy	Accuracy
S1	Daylight, normal illumination	97.0%	95.0%
S2	Night, artificial lighting	91.0%	88.0%
S3	Low-light condition	87.0%	83.0%
S4	Backlight condition	85.0%	81.0%
S5	Shadow condition	90.0%	86.0%
S6	Rainy condition	88.0%	84.0%
S7	720p camera resolution	91.0%	87.0%
S8	1080p camera resolution	97.0%	95.0%

Based on the analysis presented in Table 3, the highest system performance was achieved under normal lighting conditions, with a classification accuracy of 95.0%, whereas the lowest performance was observed under backlight conditions, with an accuracy of 81.0%. Low-light, rainy, and shadow conditions also resulted in lower performance compared with normal lighting conditions. The camera resolution evaluation showed that the 1080p camera achieved higher accuracy than the 720p camera. These results indicate that lighting conditions, weather, shadows, backlight, and camera resolution are factors that can affect the detection and classification accuracy of the system. Furthermore, a latency analysis was conducted to evaluate the system response when a potential threat was detected, using the alarm indicator as an initial security warning, as presented in Figure 12.

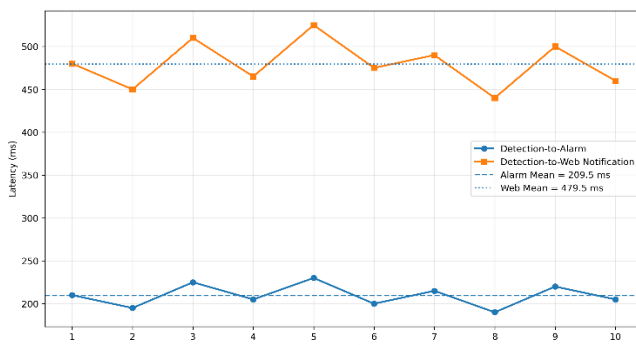


Figure 11. Detection Latency Analysis of the Early Warning System

Based on the information presented in Figure 11, the detection-to-alarm latency remained relatively stable within the range of 190–230 ms, with an average of 209.5 ms, whereas the detection-to-web notification latency ranged from 440–525 ms, with an average of 479.5 ms. These results indicate that the local alarm provided a faster response than the web notification. Furthermore, the system accuracy in detecting potentially hazardous events was evaluated using true positive (TP), true negative (TN), false positive (FP), and false negative (FN), as presented in Figure 12.

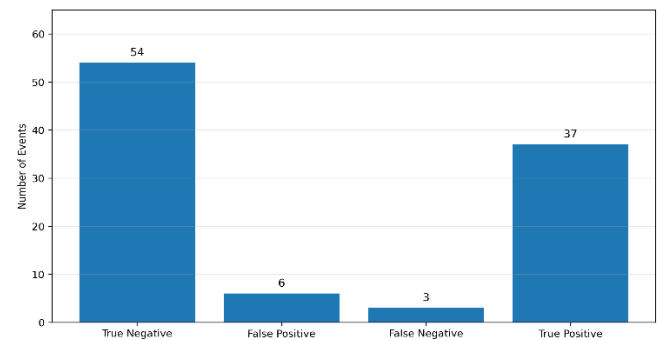


Figure 12. False Positive and False Negative Analysis

Referring to the information presented in Figure 12, the system recorded 54 true negatives, 37 true positives, 6 false positives, and 3 false negatives across 100 test events. The resulting false positive rate was 10.0%, while the false negative rate was 7.5%, indicating that a small proportion of safe interactions may still trigger alarms, while some potentially suspicious conditions may remain undetected. Finally, an evaluation of the early-warning system performance was conducted, as presented in Figure 13.

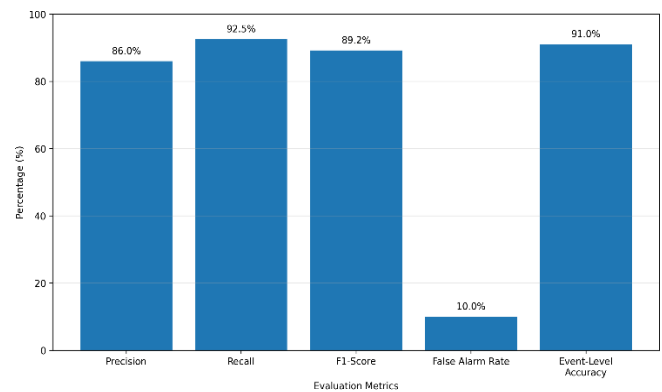


Figure 13. Performance Evaluation of the Early Warning System

Based on the analysis presented in Figure 13, the graph shows a precision of 86.0%, recall of 92.5%, F1-score of 89.2%, false alarm rate of 10.0%, and event-level detection accuracy of 91.0%. These results indicate that the system is capable of detecting most potentially suspicious events while maintaining a relatively lower false alarm rate compared with its overall detection performance.

Overall, the analyses presented in the four graphs demonstrate that the proposed system is capable of monitoring interactions between children and adults in real time through the evaluation of interaction duration, alarm frequency, and risk scores. The generated information can serve as a decision-support mechanism for identifying interactions that may potentially lead to child abduction incidents and for providing early warning notifications before more serious events occur. These findings indicate that the proposed approach has significant potential for enhancing child safety through continuous surveillance and risk-based monitoring. Based on the results of this study, several recommendations can be proposed for future research and system development. First, the proposed system can be integrated with Internet of Things technologies and advanced Artificial Intelligence techniques to further improve its detection capabilities and decision-making accuracy. The integration of these technologies is expected to enable the system to perform more intelligent contextual analysis, such as distinguishing between family members and potential abductors approaching a child. Such capabilities would significantly reduce false alarms while improving the reliability and accuracy of child abduction risk detection.

Furthermore, the proposed system can be developed into a portable and user-friendly product to facilitate practical deployment in real-world environments. With this enhancement, the child abduction early warning system could be directly implemented and utilized by the wider community as a proactive child protection solution. Ultimately, the adoption of such technologies may contribute to strengthening early child protection measures and supporting safer environments for children in residential and public areas. In addition, data privacy and security aspects need to be strengthened because the system stores captured images and information related to the potential victim and suspected perpetrator. Future development may incorporate user-based authentication and access-control mechanisms, encrypted data storage, and a data-retention policy specifying the duration for which captured images and event logs are retained before being automatically deleted. Furthermore, mechanisms for preventing unauthorized access should be implemented to ensure that the data can only be accessed by authorized personnel. These enhancements are expected to improve data security while supporting the deployment of the proposed system in real-world surveillance environments. In addition, future research should evaluate the system under various non-abduction scenarios, including parents accompanying children, adults playing with children, adults walking past children, and multiple individuals occupying the same area. Such evaluations are necessary to quantify the false alarm rate and improve the system's ability to distinguish normal social interactions from potentially suspicious conditions. The development of a context-aware risk assessment mechanism is also necessary to reduce false positives caused by normal interactions involving close proximity and prolonged interaction durations.

IV. CONCLUSION

This study developed an image-based early-warning system for identifying potentially suspicious child–adult interactions by integrating YOLOv8-based human detection, ByteTrack object tracking, *bounding-box height ratio*, *image-space distance*, and interaction duration. The experimental results demonstrate that YOLOv8 achieved an accuracy of 96.5%, precision of 97.0%, recall of 96.2%, and F1-score of 96.6%, while the proposed combination of distance and bounding-box height ratio achieved an F1-score of 91.2%, outperforming the *distance-only* (82.2%) and *height-ratio-only* (79.6%) approaches. At the event level, the system achieved a precision of 86.0%, recall of 92.5%, F1-score of 89.2%, false alarm rate of 10.0%, and event-level detection accuracy of 91.0%. The application of 3-s temporal persistence further reduced the false alarm rate to 5.0% and increased the F1-score to 93.4%, while the average detection-to-alarm latency was 209.5 ms. However, system performance was affected by camera perspective, object distance, occlusion, crowded scenes, illumination, weather conditions, and camera resolution. Therefore, the proposed system should be considered an early-warning mechanism for identifying potentially suspicious child–adult interactions rather than a definitive system for determining child abduction, with final assessment requiring further verification by an operator or authorized personnel.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Ministry of Higher Education, Science, and Technology and the BIMA Program for providing financial support through the Beginner Lecturer Research Grant (Penelitian Dosen Pemula). The authors also extend their deepest appreciation to Universitas Handayani Makassar for its institutional support and to all individuals, organizations, and community members whose contributions and assistance were invaluable to the successful completion of this research. Their support and cooperation have greatly contributed to achieving the objectives of this study.

REFERENCES

- [1] Hairullah, The Important of the Family Roles in Building the Children's Character. *Influence: International Journal Of Science Review*, 5(1), 100–106. 2023. <https://doi.org/10.54783/influencejournal.v5i1.107>.
- [2] A. Narayan, K. P. Sarang, K. Viswajith, M. R. Kowshik, P. M. Nagarajan and S. Gopalan, "Enhancing Child Safety and Security Through Integrated Technology for Urban Areas," *2024 IEEE Global Humanitarian Technology Conference (GHTC)*, Radnor, PA, USA, 2024, pp. 240–246, doi: 10.1109/GHTC62424.2024.10771557.
- [3] K. Gagneja, "Artificial Intelligence and Big Data Based Algorithm to Track Child Abduction and Child Trafficking Patterns Using Sensors," *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, Vellore, India, 2024, pp. 1–6, doi: 10.1109/ic-ETITE58242.2024.10493744.
- [4] P. S. Chandran, N. B. Byju, R. U. Deepak, K. N. Nishakumari, P. Devanand and P. M. Sasi, "Missing Child Identification System Using

- Deep Learning and Multiclass SVM," *2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, Thiruvananthapuram, India, 2018, pp. 113-116, doi: 10.1109/RAICS.2018.8635054.
- [5] L. -W. Chen, T. -P. Chen, H. -M. Chen and M. -F. Tsai, "Crowdsourced Children Monitoring and Finding With Holding Up Detection Based on Internet-of-Things Technologies," in *IEEE Sensors Journal*, vol. 19, no. 24, pp. 12407-12417, 15 Dec.15, 2019, doi: 10.1109/JSEN.2019.2938324.
- [6] B. Vinavatani, M. R. Panna, P. H. Singha and G. J. W. Kathrine, "AI for Detection of Missing Person," *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India, 2022, pp. 66-73, doi: 10.1109/ICAAIC53929.2022.9792672.
- [7] F.Gerry, J.Joniwan, W.Agung, and T.Budi. Preventing Child Kidnaping at Home Using CCTV that Utilizes Face Recognition with You Only Look Once (YOLO) Algorithm. *Journal of Social Research*. 2. 3291-3304, 2023. DOI: 10.55324/josr.v2i9.1403.
- [8] S.Jura, A.Jalil, Abdul, H.Jusman, and M.Firman. Integrasi Sensor Multi-Modal pada Ikat Pinggang Smart sebagai Perlindungan Penculikan Anak. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*. 12. 321. 2024. DOI: 10.26760/elkomika.v12i2.321.
- [9] P. Agrawal *et al.*, "YOLO Algorithm Implementation for Real Time Object Detection and Tracking," *2022 IEEE Students Conference on Engineering and Systems (SCES)*, Prayagraj, India, 2022, pp. 01-06, doi: 10.1109/SCES55490.2022.9887678.
- [10] A. Odeh and N. Odeh, "OpenCV and its Applications in Artificial Intelligent Systems," *2024 International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS)*, Dubrovnik, Croatia, 2024, pp. 242-249, doi: 10.1109/ICCNS62192.2024.10776047.
- [11] H. A. A. Putra, A. Murni Arymurthy and D. Chahyati, "Enhancing Bounding Box Regression for Object Detection: Dimensional Angle Precision IoU-Loss," in *IEEE Access*, vol. 13, pp. 81029-81047, 2025, doi: 10.1109/ACCESS.2025.3567767.
- [12] Y. Wang and V. Y. Mariano, "A Multi Object Tracking Framework Based on YOLOv8s and Bytetrack Algorithm," in *IEEE Access*, vol. 12, pp. 120711-120719, 2024, doi: 10.1109/ACCESS.2024.3450370.