A Lightweight Deep Learning Approach for Autonomous Detection of Risky Infant Sleep Postures

Gabriel L. Pirotello, Ademar T. Akabane, Diana C. González

[1]Polytechnic School – Pontifical Catholic University of Campinas, Brazil

gabriel.lp3@puccampinas.edu.br

{ademar.akabane,diana.cristina}@puc-campinas.edu.br

Abstract. *Sudden Infant Death Syndrome (SIDS) and sleep-related fatalities remain a leading cause of post-neonatal mortality, often linked to hazardous sleep positions such as prone positioning and airway obstruction. While continuous monitoring is essential for prevention, existing wearable sensors pose risks of skin irritation and displacement, while traditional video monitors depend on fallible human vigilance. This study addresses these limitations by developing a high-performance, non-invasive vision-based system for the autonomous detection of risky infant postures. Utilizing the YOLOv8n architecture, we implemented a lightweight framework optimized for real-time classification of "Safe" and "Danger" positions. Our findings demonstrate high detection reliability, with the model achieving a recall of 95.11% and a precision of 96.07% for hazardous postures, ensuring a critical reduction in false negatives while maintaining caregiver trust. To mitigate the "black-box" nature of deep learning, we applied Gradient-weighted Class Activation Mapping (Grad-CAM). Qualitative analysis confirmed that the model's predictive logic is grounded in clinically relevant anatomical features, such as the infant's head and torso, rather than environmental noise. These results establish a robust computational foundation for future privacy-preserving, localized monitoring solutions that can provide immediate, life-saving alerts in home environments.*

Keywords: Infant Posture Detection, Computer Vision, YOLOv8, Risk Detection, SIDS, Safe Sleep

1. Introduction

The promotion of safe sleep environments is a cornerstone of pediatric preventative care, primarily aimed at reducing the incidence of Sudden Infant Death Syndrome (SIDS) and other sleep-related fatalities [Bharati 2021, Tan and Goh 2024, Abdelfattah et al. 2025, Udoko et al. 2026]. Despite widespread public health campaigns such as the "Back to Sleep" initiative, hazardous sleeping conditions, including prone positioning, face occlusion, and airway obstruction, remain a leading cause of post-neonatal mortality globally [Ogbo et al. 2019, Izulla et al. 2023, Maugeri et al. 2025]. Current pediatric guidelines, most notably from the American Academy of Pediatrics [Moon et al. 2016, Yamada et al. 2024], emphasize the necessity of supine positioning to maintain clear airways. However, continuous manual supervision by caregivers is often impractical, particularly during nocturnal periods, creating a critical need for automated, reliable, and non-invasive monitoring solutions [Tan and Goh 2024, Maugeri et al. 2025, Abdelfattah et al. 2025].

Existing infant monitoring technologies have traditionally relied on wearable sensors [Cay et al. 2022, Chandnani et al. 2025] or basic video baby monitors [Alam et al. 2023]. While wearables can track physiological vitals, they are often prone to displacement, cause skin irritation, and may pose entanglement risks [Tan and Goh 2024]. Conversely, standard video monitors require constant human vigilance, which is susceptible to caregiver fatigue and reduced system trust [Abdelfattah et al. 2025]. Recent advancements in Deep Learning and Computer Vision have enabled automated posture detection. Nonetheless, significant gaps remain: many existing models lack the computational efficiency required for real-time inference or fail to provide the interpretability required for safety-critical healthcare applications.

Recent studies have explored infant sleep posture monitoring across diverse sensing modalities [Yun et al. 2020, Cay et al. 2022, Alam et al. 2023, Chandnani et al. 2025]. Traditional technologies primarily rely on wearable sensors or conventional video monitors. Wearable devices can track physiological signals and body motion but are often affected by displacement, skin irritation, and potential entanglement risks [Tan and Goh 2024]. Conventional video systems, in turn, demand continuous caregiver supervision, leading to fatigue and inconsistent monitoring reliability [Abdelfattah et al. 2025]. Sensor fusion approaches integrate acceleration, ECG, and pressure to enhance sleep posture classification [Wu et al. 2025, Li et al. 2025]. Although these methods can achieve high accuracy, they typically require intrusive hardware, such as instrumented beds or wearable devices, limiting their applicability in unobtrusive home monitoring scenarios. Recent advances in deep learning and computer vision have enabled contactless, image-based posture recognition [Bharati 2021]. More recently, Huang and Huang [Huang and Huang 2024] proposed a hybrid infant posture recognition framework that combines YOLOv5, OpenPose, and SVM-based feature fusion, reporting improved classification performance compared to alternative detection architectures such as SSD and Faster R-CNN. Nevertheless, significant challenges remain: many existing solutions lack the computational efficiency required for real-time edge deployment or fail to provide the interpretability essential for safety-critical pediatric healthcare applications.

The purpose of this work is to address these gaps by developing and validating a high-performance vision-based system for the autonomous detection of risky infant postures. Utilizing the YOLOv8n (nano) architecture, we propose a lightweight framework optimized for high-sensitivity classification of *Safe* and *Danger* sleep positions. Recent studies have demonstrated the suitability of YOLO-based architectures for infant posture recognition and real-time monitoring applications due to their favorable balance between detection accuracy and computational efficiency [Ge et al. 2021, Huang and Huang 2024, Pereira 2025]. Building upon these findings, the proposed framework adopts an end-to-end lightweight design focused on edge-oriented deployment and safety-critical monitoring reliability. Our findings demonstrate that through iterative dataset refinement and a focus on visual diversity, the model achieves a recall of 95.11% and a precision of 96.07% for hazardous postures. Furthermore, by employing Gradient-weighted Class Activation Mapping (Grad-CAM), we provide qualitative evidence that the model's predictive logic is grounded in clinically relevant anatomical features, such as the infant's head and torso, rather than environmental noise.

This research offers some primary contributions to the field of intelligent health monitoring:

- The curation of a balanced dataset comprising 7,870 images reflects diverse real-world conditions, including variations in lighting, clothing, and camera angles;
- The implementation of a YOLOv8n-based model that prioritizes the detection of high-risk scenarios (Recall) to minimize dangerous false negatives;
- The application of Grad-CAM to validate that the neural network focuses on

anatomically significant regions, thereby mitigating the "black-box" nature of deep learning in a safety-critical context.

The remainder of this paper is structured as follows. Section 2 details the proposed methodology. Section 3 analyzes the experimental results obtained from the proposed solution. Finally, Section 4 summarizes the key findings of the study.

2. Methodology

The methodology follows a four-pillar framework: dataset engineering (Section 2.1), architectural implementation (Section 2.2), performance evaluation (Section 2.3), and interpretability analysis (Section 2.4). The study utilized an iteratively expanded dataset of 7,870 images to train a YOLOv8n model, chosen for its real-time inference efficiency. System performance was quantified using precision and recall to prioritize the detection of hazardous postures, while Grad-CAM was employed to ensure the model's transparency by validating its focus on clinically relevant anatomical features.

2.1. Dataset Preparation

The dataset was manually curated to represent a broad spectrum of infant sleep scenarios, encompassing both recommended safe conditions and potentially hazardous ones. It was structured to support a binary classification task with two classes: *Safe* and *Danger*. Sample annotation was performed manually by trained annotators based on visual indicators, including face visibility, body orientation, and risk of airway obstruction. Labeling criteria were defined in accordance with established pediatric sleep safety guidelines, specifically the recommendations of the American Academy of Pediatrics [Moon et al. 2016, Yamada et al. 2024]. For ambiguous cases, annotators followed a conservative decision protocol grounded in the asymmetric clinical cost of misclassification. Lateral postures with partial face visibility were assigned to the *Safe* class only when two conditions were simultaneously met: clear evidence of an unobstructed airway and a stable lateral recovery position that presented no biomechanical tendency toward prone rollover. Conversely, lateral postures exhibiting progressive facial occlusion, restricted upper-limb mobility, or a postural configuration consistent with an increased risk of transitioning to the prone position were assigned to the *Danger* class, regardless of whether full airway obstruction had yet occurred. This conservative labeling policy prioritizes sensitivity to emerging risk over specificity, reflecting the clinical principle that the cost of a missed hazardous event substantially outweighs the cost of a false alarm in infant sleep monitoring contexts.

The initial dataset comprised approximately 2,000 images. To improve class balance and clinical representativeness, four iterative rounds of targeted data acquisition were conducted. Each round focused on supplementing underrepresented hazardous conditions, including prone positioning, facial occlusion, and restricted limb movement, to enrich the *Danger* class and broaden the diversity of risk-related scenarios. Following these refinement cycles, the final dataset comprised 7,870 images, distributed across classes at approximately 59.32% *Safe* and 40.68% *Danger*. This distribution reflects a deliberate trade-off between two competing requirements: ensuring sufficient representational coverage of critical risk scenarios, and preserving the broader visual diversity inherent to safe sleeping postures. The dataset was partitioned into training (70%), validation (20%), and test (10%) subsets, with strict non-overlapping splits enforced across all three sets to prevent data leakage and ensure unbiased evaluation.

2.2. Model Architecture, Training, and Preprocessing

The YOLOv8n (nano) architecture was selected as the primary object detection framework due to its optimized trade-off between diagnostic accuracy and the low-latency computational efficiency required for real-time edge inference. The model implementation and training protocols were executed as follows:

- Architectural Framework: The model was developed using the Ultralytics YOLOv8 framework, implemented in PyTorch. This lightweight variant is specifically designed to minimize the memory footprint while maintaining high sensitivity in safety-critical classification tasks;
- Transfer Learning and Initialization: To accelerate convergence and utilize generalized visual hierarchies, weights were initialized from the COCO dataset. Fine-tuning was subsequently conducted on the infant-specific dataset for 200 epochs;
- Hyperparameter Configuration: Training was performed with an input resolution of 640×640 pixels and a batch size of 16. The Adam optimizer was utilized with a fixed learning rate of 10^{-3}, ensuring stable gradient descent without the use of a learning rate scheduler;
- Loss Function and Optimization: The network was optimized using a composite loss function that integrates bounding box regression (box-loss), classification (cls-loss), and distribution focal loss (dfl-loss). GPU acceleration was employed to enhance training stability and convergence speed;
- Preprocessing and Augmentation Policy: Experimental evaluation of standard data augmentation techniques, including rotations and brightness perturbations, revealed a degradation in performance for the Danger class. Consequently, no additional transformations were applied to the training pipeline. Images derived from public repositories that already contained basic adjustments (e.g., contrast enhancement or cropping) were retained to maintain visual consistency across the dataset;
- Validation and Selection: Performance was monitored continuously using precision, recall, and mAP@0.5 on a dedicated validation set. Model checkpoints were generated every 25 epochs, with the final weights selected based on the optimal mAP@0.5 configuration to ensure maximum generalizability.

2.3. Performance Evaluation

The efficacy of the detection model was quantified using standard object detection metrics derived from the intersection over union (IoU) between predicted and ground-truth bounding boxes. Given the high-stakes nature of infant monitoring, the evaluation framework was designed to prioritize sensitivity to high-risk scenarios while maintaining overall system precision.

1. Quantitative Indicators: Performance was assessed using Precision, Recall, and mean Average Precision (mAP) at an IoU threshold of 0.5 (mAP@0.5). To evaluate the model's localization robustness across varying overlap constraints, the mAP@0.5:0.95 metric was additionally computed;
2. Safety-Critical Prioritization: In this application, Recall for the Danger class served as the primary performance indicator, as the failure to detect a hazardous posture (False Negative) poses a significant risk to infant safety. Conversely, Precision was monitored to minimize the incidence of false alarms, which could lead to caregiver alarm fatigue and reduced system trust;

3. Error Characterization: A Confusion Matrix was generated to provide a granular analysis of misclassifications between the Safe and Danger categories. This diagnostic tool allowed for the identification of specific visual edge cases where the model might exhibit classification bias.
4. Model Interpretability: To ensure that detections were based on clinically relevant features rather than background artifacts, Grad-CAM was utilized. This technique produced visual heatmaps highlighting the specific regions, such as the infant's torso or head, that exerted the strongest influence on the network's predictive decisions.

2.4. Model Interpretability

To mitigate the "black-box" nature of deep neural networks and ensure their suitability for safety-critical healthcare monitoring, an interpretability analysis was conducted using Grad-CAM. This post-hoc explanation technique provides qualitative evidence of the model's decision-making process by highlighting the regions that most strongly influence its predictions.

- Mechanism of Attribution: Grad-CAM utilizes the gradient information flowing into the final convolutional layer of the YOLOv8n backbone to produce a localization map. These gradients are globally averaged to compute importance weights for each feature map, which are subsequently combined through a ReLU activation to highlight pixels with a positive influence on the predicted class;
- Clinical Validation: The resulting heatmaps were overlaid on test-set frames to verify that the model's attention remains focused on anatomically relevant regions, such as the infant's head, torso, and limb orientation;
- Spurious Correlation Filtering: A primary objective of this analysis was to confirm that the system ignores non-contributory background artifacts—such as crib railings, bedding textures, or environmental sensors—thereby reducing the risk of predictions based on environmental bias rather than posture-specific indicators;
- Temporal Consistency: Activation maps were extracted across sequential video frames to evaluate the stability of the model's focus during infant movement, providing a robust baseline for the system's reliability in dynamic, real-world home settings.

Activation maps were generated from representative frames extracted from the test dataset and overlaid on the corresponding original images to visually inspect the model's attention patterns. The main findings of this analysis are presented in Section 3.

3. Results and Discussion

The proposed YOLOv8n-based system was comprehensively evaluated through quantitative benchmarking, confusion matrix analysis, and Grad-CAM interpretability. Together, these assessments establish the model's suitability for deployment in safety-critical, real-time infant monitoring contexts. The evaluation framework was explicitly designed to prioritize Recall for the Danger class, in recognition of the asymmetric cost of false negatives in this domain: an undetected hazardous posture may directly contribute to airway obstruction and elevated SIDS risk, while a false alarm, though disruptive, does not pose an immediate physiological threat.

3.1. Impact of Iterative Dataset Refinement on Model Performance

A primary methodological approach of this study is the demonstration that targeted, iterative dataset expansion yields consistent and measurable gains in detection performance, particularly for safety-critical minority classes. The training dataset was expanded across five discrete stages from an initial 2,000-image baseline to a final cohort of 7,870 images, with each increment of approximately 1,000 images prioritizing the acquisition of clinically representative Danger samples, specifically prone positioning and face occlusion. Figure 1 documents the evolution of four canonical object detection metrics at each stage.

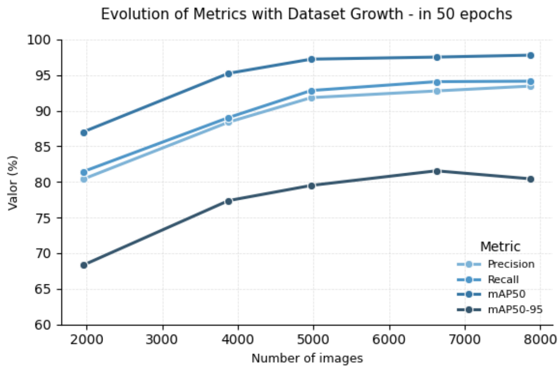

Figure 1. Performance metrics as the dataset size increased.

The most diagnostically significant trend is the monotonic improvement in Danger class Recall, which increased from 81.45% at the 2,000-image baseline to 94.14% at the penultimate stage, and achieved 95.11% upon final model training over 200 epochs. This represents an absolute gain of 13.66 percentage points relative to the initial configuration. The mAP@0.5 metric, which integrates precision and recall across the full detection confidence range, similarly stabilized above 97% at later stages, indicating that performance gains were not achieved at the expense of localization accuracy.

From a statistical learning perspective, these results offer empirical evidence for two well-established phenomena in deep learning for medical imaging. First, class imbalance sensitivity: the Danger class, representing 40.68% of the final dataset, is intrinsically harder to generalize without sufficient intra-class visual diversity. Prone positioning, in particular, encompasses a wide range of appearances depending on infant age, clothing opacity, and camera angle, making early models prone to under-sampling this phenotypic variation. Second, feature distribution coverage: as noted in analogous computer vision studies on neonatal monitoring [Bharati 2021, Cay et al. 2022], models trained on narrow visual distributions exhibit a systematic performance degradation on out-of-distribution test examples. The iterative expansion strategy directly addresses this limitation by explicitly seeking samples that populate underrepresented regions of the feature space.

Importantly, the mAP@0.5:0.95 metric, which averages precision over IoU thresholds from 0.50 to 0.95 and thus penalizes imprecise bounding box localization, also demonstrated steady improvement across dataset expansion stages, rising from approximately 68% at the initial stage to 83.19% at the final configuration. This indicates that dataset growth improved not only classification confidence but also spatial localization fidelity, a finding that has direct relevance for potential downstream applications requiring precise region-of-interest cropping for secondary anatomical analyses.

Table 1 summarizes the final performance metrics. The model achieved a precision of 96.07% and a recall of 95.11% for the *Danger* class, along with an overall mAP@0.5 of 97.12% and mAP@0.5:0.95 of 83.19%. These results demonstrate high robustness and confirm the model's ability to detect potentially hazardous postures across a range of environmental and visual conditions. The high recall helps reduce the likelihood of missed detections, critical in safety-sensitive scenarios, while the high precision minimizes false alarms, preserving caregiver trust in the system.

Table 1. Final Performance Metrics of the Model Trained for 200 Epochs.

Metric	Overall	Danger Class	Safe Class
Precision (%)	96.40	96.07	97.68
Recall (%)	95.52	95.11	95.93
mAP@0.5 (%)	97.12	96.50	97.75
mAP@0.5:0.95 (%)	83.19	83.86	82.53

3.2. Training Dynamics and Generalization Evidence

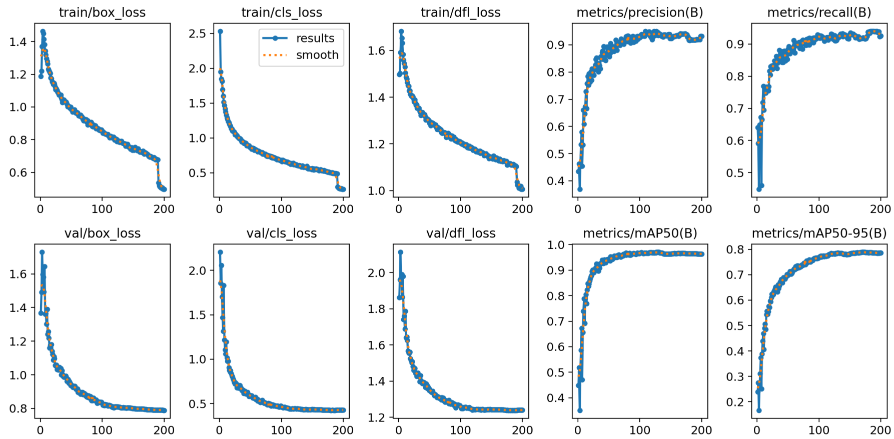

Figure 2. Training and validation curves over 200 epochs.

Figure 2 presents the full 200-epoch training trajectory, plotting the co-evolution of three composite loss components (bounding box regression loss: `box_loss`; classification loss: `cls_loss`; and distribution focal loss: `dfl_loss`) alongside the four principal performance metrics for both the training and validation partitions. Several features of this trajectory warrant explicit discussion. First, all three loss components demonstrate smooth, monotonically decreasing convergence without evidence of instability or oscillatory behavior across the full training horizon. This smooth convergence profile is attributable in part to the composite loss formulation inherent to the YOLOv8 architecture, which simultaneously optimizes spatial localization and categorical discrimination.

Second, and critically for generalizability assessment, the training and validation loss curves remain closely aligned throughout the 200-epoch training run, with no observable divergence. Performance metrics, including precision, recall, and mAP@0.5, stabilized after approximately 150 epochs on the validation set, with no further statistically meaningful gains in the final 50 epochs. This plateau behavior indicates that the model reached a stable local optimum within the loss landscape and was not simply memorizing training-set idiosyncrasies. The use of COCO-pretrained weight initialization likely contributed to this rapid and stable convergence by supplying a rich ensemble of low- and mid-level visual feature detectors that transferred effectively to the infant posture domain.

Third, the deliberate decision to omit standard data augmentation (rotations, brightness perturbations, and horizontal flips) deserves technical elaboration. Empirical ablation experiments conducted during pilot training revealed that aggressive geometric augmentation paradoxically degraded Danger class Recall. This finding is consistent with the hypothesis that spatial orientation is a semantically load-bearing feature for posture classification: rotating a prone infant image by 90° or 180° fundamentally alters the gravitational semantics of the posture, potentially training the model to classify physically impossible or clinically nonsensical configurations.

Figure 3. Example of real-time inference showing bounding boxes and class labels.

Figure 3 illustrates a representative real-time inference example in which the model simultaneously detects and classifies two infants occupying distinct regions of the camera field. The left infant is assigned a *Danger* label with a confidence score of 0.94, while the right infant receives a *Safe* classification with a score of 0.88. The scene involves non-uniform illumination and a moderately oblique camera angle, conditions that constitute known stress factors for object detection models. The accurate, high-confidence dual detection in this challenging configuration demonstrates the model's practical viability in realistic home environments, where multiple infants sharing a sleep space, variable lighting from adjacent rooms, and sub-optimal camera placement are common.

To further evaluate the practical applicability of the proposed framework, inference performance was also analyzed. The YOLOv8n-based model achieved an average inference time of 72.56 ms per frame, corresponding to approximately 21.28 FPS. These results demonstrate that the proposed solution is compatible with real-time monitoring requirements and reinforce its suitability for lightweight edge-oriented deployment scenarios.

3.3. Error Analysis and Clinical Risk Characterization

A confusion matrix was generated from the test set to characterize the nature and distribution of classification errors. Figure 4 presents the full matrix with derived class-level recall and overall accuracy. The system correctly classified 254 of 260 Danger instances (97.69% class-level recall) and 592 of 603 Safe instances (98.18% class-level recall), yielding an overall accuracy of 98.03% (846 of 863 test samples). The total error budget consists of 6 false negatives (FN) and 11 false positives (FP). From a clinical risk perspective, these two error categories carry markedly asymmetric consequences.

The 6 false negatives, instances where a hazardous posture was classified as Safe, represent the most clinically consequential error type. Qualitative examination of these cases suggests that misclassifications arise predominantly in visually ambiguous borderline scenarios: near-lateral postures that approach the prone threshold without fully obstructing the airway, or infants whose face remains partially visible despite a technically compromised respiratory position. The 11 false positives, Safe instances misclassified as Danger, are proportionally distributed across scenarios involving atypical but non-

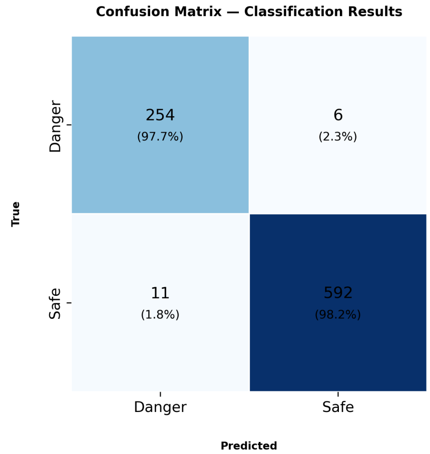

Figure 4. Confusion matrix generated on the test set.

hazardous postures: lateral positioning with partial face occlusion, or safe prone positioning of infants old enough to independently reposition (a class boundary not encoded in the current binary labeling scheme).

3.4. Model Interpretability: Grad-CAM Activation Analysis

Figure 5 presents a mosaic of Grad-CAM overlays extracted from sequential video frames, including both Danger and Safe classification examples. Three substantive observations emerge from this analysis. First, across all sampled frames, the model's peak activation consistently coincides with the infant's head and upper torso region, the anatomical substrates directly implicated in airway obstruction risk. This focus is clinically appropriate: the orientation and visibility of the infant's face and neck are the primary determinants of respiratory patency, and the model has, without explicit anatomical supervision, learned to attend to these structures. Second, the activation maps exhibit negligible response to background elements, including crib railings, bedding textures, ambient lighting gradients, and environmental sensors, a finding that substantially reduces concern about spurious environmental confounders driving predictions. Third, the activation patterns demonstrate temporal-spatial consistency across consecutive frames, suggesting that the model's attention is anchored to genuine anatomical landmarks rather than frame-level artifacts or compression noise, an important property for robust real-world deployment in streaming video inference pipelines.

Collectively, the Grad-CAM evidence satisfies two independent validation criteria. From a technical perspective, it confirms that the model has learned semantically grounded representations rather than dataset-specific shortcuts. From a clinical perspective, it provides qualitative evidence to end-users and regulatory evaluators that the system's predictions can be meaningfully interpreted with reference to observable anatomical features.

4. Conclusion

The mitigation of Sudden Infant Death Syndrome remains a critical challenge in pediatric medicine, necessitating the development of robust, non-invasive, and autonomous monitoring solutions. This study engineered a high-performance computer vision framework based on the YOLOv8n architecture, specifically optimized for the real-time classification of hazardous infant sleep postures. By curating a diverse dataset of 7,870 images

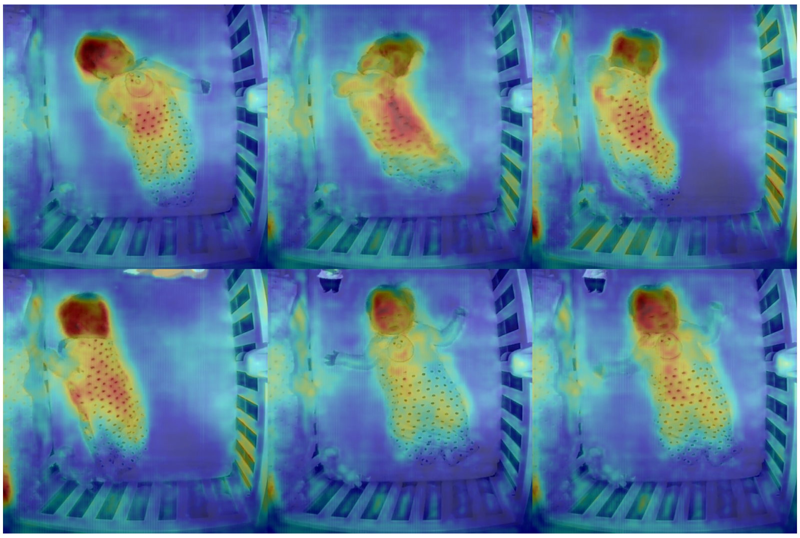

Figure 5. Activation regions of the model in frames extracted from a video.

through iterative refinement, we achieved a significant recall of 95.11% for danger-prone scenarios. This high sensitivity to risk, coupled with a precision of 96.07%, effectively addresses the dual challenges of ensuring infant safety and preventing caregiver alarm fatigue. Beyond quantitative performance, this research advances the interpretability of deep learning in safety-critical healthcare. Through Grad-CAM analysis, we provided qualitative evidence that the model's predictive logic is grounded in clinically relevant anatomical features, such as head and torso orientation, rather than environmental artifacts. This transparency is essential for bridging the gap between automated detection systems and clinical trust.

Ultimately, this work establishes a scalable computational foundation for the next generation of intelligent neonatal care. While currently validated as a standalone vision system, the proposed architecture is designed for seamless integration into localized, privacy-preserving monitoring ecosystems. Future efforts will focus on multimodal sensor fusion to incorporate environmental and physiological data, moving toward a proactive, predictive alerting model capable of intervening before hazardous events occur.

References

Abdelfattah, M., Zhou, L., Sum-Ping, O., Hekmat, A., Galati, J., Gupta, N., Adaimi, G., Marwaha, S., Parekh, A., Mignot, E., et al. (2025). Automated detection of isolated rem sleep behavior disorder using computer vision. *Annals of Neurology*, 97(5):860–872.

Alam, H., Burhan, M., Gillani, A., Haq, I. U., Arshed, M. A., Shafi, M., and Ahmad, S. (2023). Iot based smart baby monitoring system with emotion recognition using machine learning. *Wireless Communications and Mobile Computing*, 2023(1):1175450.

Bharati, V. (2021). An efficient edge deep learning computer vision system to prevent sudden infant death syndrome. In *2021 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 286–291. IEEE.

Cay, G., Solanki, D., Al Rumon, M. A., Ravichandran, V., Hoffman, L., Laptook, A., Padbury, J., Salisbury, A. L., and Mankodiya, K. (2022). Neowear: An iot-connected e-textile wearable for neonatal medical monitoring. *Pervasive and Mobile Computing*, 86:101679.

Chandnani, K., Tripathy, S., Parbhakar, A. K., Takiar, K., Singhal, U., Sasikumar, P., and Maheswari, S. (2025). A novel smart baby cradle system utilizing iot sensors and machine learning for optimized parental care. *Scientific Reports*, 15(1):19080.

Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J. (2021). Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430.*

Huang, X. and Huang, M. (2024). Classification and recognition of infant sleeping positions based on dual model feature fusion. In *Proc. IEEE Int. Conf. on Artificial Intelligence, Automation and High Performance Computing (AIAHPC)*, pages 57–63.

Izulla, P., Muriuki, A., Kiragu, M., Yahner, M., Fonner, V., Nitu, S. N. A., Osir, B., Bello, F., and de Graft-Johnson, J. (2023). Proximate and distant determinants of maternal and neonatal mortality in the postnatal period: A scoping review of data from low-and middle-income countries. *PLoS One*, 18(11):293–479.

Li, C., Ren, G., and Wang, Z. (2025). Sleep posture recognition method based on sparse body pressure features. *Applied Sciences*, 15(9).

Maugeri, A., Barchitta, M., Schillaci, G., and Agodi, A. (2025). Spatial patterns and temporal trends in stillbirth, neonatal, and infant mortality: an exploration of country-level data from 2000 to 2021. *Journal of Global Health*, 15:04034.

Moon, R. Y., Darnall, R. A., Feldman-Winter, L., Goodstein, M. H., and Hauck, F. R. (2016). SIDS and other sleep-related infant deaths: Updated 2016 recommendations for a safe infant sleeping environment. *Pediatrics*, 138(5):e20162938.

Ogbo, F. A., Ezeh, O. K., Awosemo, A. O., Ifegwu, I. K., Tan, L., Jessa, E., Charwe, D., and Agho, K. E. (2019). Determinants of trends in neonatal, post-neonatal, infant, child and under-five mortalities in tanzania from 2004 to 2016. *BMC public health*, 19(1):1243.

Pereira, A. G. (2025). Uma abordagem leve para detecção de embarcações utilizando redes neurais sem pesos e regressão. Dissertação de mestrado, Universidade Federal do Rio de Janeiro (UFRJ), COPPE, Programa de Engenharia de Sistemas e Computação, Rio de Janeiro, RJ, Brasil. Acesso em: 7 maio 2026.

Tan, J. H. and Goh, C. P. (2024). Enhancing child safety: Computer vision-based accident detection for infants and toddlers. In *2024 3rd International Conference on Digital Transformation and Applications (ICDXA)*, pages 1–5. IEEE.

Udoko, A. N., Dyess, N. F., Hwang, S. S., and Fisher, C. R. (2026). Sudden unexpected infant death and safe sleep practices. *NeoReviews*, 27(1):e1–e10.

Wu, L., Guo, S., Han, L., and Jia, J. (2025). Human sleep motion recognition based on multi-sensor measurement data fusion, measurement. *Measurement*, 253, Parte C(1):117746.

Yamada, N. K., Szyld, E., Strand, M. L., Finan, E., Illuzzi, J. L., Kamath-Rayne, B. D., Kapadia, V. S., Niermeyer, S., Schmölzer, G. M., Williams, A., et al. (2024). 2023 american heart association and american academy of pediatrics focused update on neonatal resuscitation: an update to the american heart association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. *Circulation*, 149(1):157–166.

Yun, I., Jeung, J., Kim, M., Kim, Y.-S., and Chung, Y. (2020). Ultra-low power wearable infant sleep position sensor. *Sensors*, 20(1).