A Deep Learning Framework for Infant Sleep Monitoring and SIDS Risk Assessment

Abhishek Herbert Samuel[1], Kakaraparty Sashank Kumar[1], Aman Bhat[1], Parthiv Susarla[1], Ashwinth Janarthanan[2], Prakash Poornachari[3], and Vidhusha Srinivasan[1]

[1] Immersive Technologies Lab, Department Of Computer Science And Engineering, School Of Engineering, Shiv Nadar University, Chennai, India
[2] Meta Learning Research Laboratory, Department of Computer Applications, National Institute of Technology, Tiruchirappalli, India
[3] Department of Electronics Engineering, MIT Campus, Anna University, Chennai, India

Abstract. Sudden Infant Death Syndrome (SIDS) is a major concern in infant care. The infant's sleeping position is one of the key risk factors. Most monitoring systems today use sensors to track physiological signals. While these systems can work well in controlled environments, they are not always comfortable for long-term use and can be difficult to use regularly at home.

In this work, we developed a vision-based approach using a deep learning model to classify an infant's sleeping position. The model has been trained on a dataset of more than 3,000 infant sleep images which were divided into safe and unsafe categories. A transfer learning method was used for this binary classification task.

The model performed well on the test dataset, with an accuracy of 97.97% and a ROC-AUC of 0.9964. However, when testing was done in real-world conditions, the results were not always consistent. During video-based testing, especially with low-quality input and during deployment on an NVIDIA Jetson Nano, some misclassifications were observed. In a few cases, the model produced incorrect predictions even when the posture was clearly identifiable.

Lighting, background and overall visual conditions such as camera setup affected the predictions more than expected. To improve this, more diverse training data will be needed, especially real-world video, along with methods that help the model handle such conditions better.

An exploratory temporal consistency-based correction experiment inspired by concepts associated with Physics-Informed Neural Networks (PINNs) was therefore performed to see if isolated false positive predictions can be corrected.

Keywords: Infant Sleep Monitoring · SIDS Risk Assessment · Real-World Deployment · Deployment Challenges · Video-Based Evaluation · Deep Learning · Physics-Informed Neural Networks

1 Introduction

Unexplained infant deaths, especially during sleep are primarily caused by Sudden Infant Death Syndrome (SIDS) [1, 2]. Evidence available from the medical field shows that sleep posture is a major risk factor with sleeping in the supine position being the safer option compared to the risky prone position [1]. Therefore, the need for sleep safety practices is important, but in real-world scenarios, it is not feasible for caregivers to constantly monitor sleep manually.

To solve this problem, previous works have focused on technology-based monitoring systems, in particular the use of Internet of Things (IoT) frameworks for continuous monitoring and alert generation [3–5]. These systems allow monitoring of physiological parameters and environmental conditions, but often depend on contact-based sensors, which are invasive and prone to placement, motion artifacts and signal instability.

Recent advances in computer vision and deep learning provide a non-invasive alternative by enabling posture recognition directly from visual data. Convolutional Neural Networks have demonstrated strong performance in healthcare-related classification tasks, including in-bed posture detection [6]. Such approaches offer the potential for scalable and unobtrusive infant monitoring systems without requiring physical contact.

However, these approaches primarily emphasize posture classification under controlled evaluation settings while the model's performance under practical operating conditions remains comparatively less explored.

In this work, we present a deep learning-based approach for infant sleep monitoring using a transfer learning framework. The model is trained on a curated dataset of infant sleep images and designed to distinguish between safe (supine) and unsafe (prone or obstructed) conditions based on visual features.

In addition to image-based evaluation, the proposed system is examined under deployment-oriented conditions through real-time video processing, including scenarios with low-resolution and low-quality input, and preliminary hardware implementation on an edge device (NVIDIA Jetson Nano) using a cost-effective camera-based pipeline. This enables assessment of model behavior beyond controlled datasets in more realistic operating conditions.

While the model achieves strong performance under image-based evaluation, observations from video-based processing and hardware deployment indicate variations in prediction behavior when applied to deployment-like environments. These variations are influenced by differences in visual characteristics and scene context between training data and real-world conditions, highlighting challenges in generalization.

The main contributions of this work are as follows:

- A deep learning-based sleep monitoring system to detect unsafe and safe positions in infant sleep.
- Development of a curated infant sleep image dataset for model training, validation and testing.
- Real time model performance evaluation done, even with low-quality and low-resolution images.

- Analysis of hardware and software-based implementations in order to understand the gap between model performance and real-world conditions.
- Exploratory observations on temporal consistency-based correction mechanisms in order to reduce isolated false positive predictions during deployment.

2 Related Work

Existing approaches to reducing the risk of Sudden Infant Death Syndrome (SIDS) have predominantly focused on physiological monitoring using sensor-based systems. Prior works have explored Internet of Things (IoT)-based architectures for continuous monitoring of vital parameters such as heart rate and respiration rate [3–5, 7]. These systems enable real-time observation and can generate alerts when abnormal conditions are detected.

While effective in controlled settings, such approaches rely heavily on contact-based sensors attached to the infant's body, making them inherently intrusive. Their performance is sensitive to sensor placement, motion artifacts and signal instability, which may lead to inaccuracies during continuous monitoring. Additionally, dependence on multiple hardware components increases the system's complexity and limits ease of deployment in real-world home environments [8, 9].

To address these limitations, recent advancements in computer vision and deep learning have introduced non-invasive alternatives for infant monitoring. Vision-based systems analyze image data to infer posture and environmental conditions without requiring physical contact. Convolutional neural networks have demonstrated strong performance in posture classification tasks, including in-bed posture recognition [6]. These approaches offer a scalable and unobtrusive solution for monitoring infant sleep conditions.

However, most vision-based posture classification studies are evaluated on curated image datasets and do not explicitly consider deployment constraints such as real-time processing, deployment on an edge device and variability in real-world environments [3]. In particular, domain differences between training data and deployment scenarios such as variations in lighting, background and representation of the subject can significantly impact model performance.

Building upon these observations, this study focuses not only on classification performance but also on evaluating model behavior when deployed in the real-world. The proposed approach investigates the impact of domain differences through video-based evaluation and preliminary hardware deployment, providing insights into the challenges of translating image-trained models to real-world infant monitoring systems. Recent studies have also highlighted the role of domain generalization and representation learning in improving robustness of deep learning models under varying deployment conditions [10].

3 Proposed Methodology

This section describes the methodology adopted for classifying the infant's sleeping position. The overall approach consists of dataset preparation, model development and the design of the system pipeline. Each component is detailed in the following subsections.

3.1 Dataset

The dataset used in this study consists of images of the infant's sleeping position. It is categorized into two classes: *safe* and *unsafe*. Safe images correspond to supine sleeping positions. Unsafe images include prone positions and positions where there are objects on the crib or the infant's body, particularly its face.

The dataset was curated from multiple publicly available online image sources and subsequently preprocessed to ensure quality and consistency. Unreadable and corrupted images were removed through validation using both the Pillow library (PIL) and TensorFlow decoding pipelines. Duplicate images were identified and eliminated to prevent data leakage and bias during training.

During preprocessing, images were inspected to ensure that the infant posture was clearly visible and that the labels matched the intended class. Images with poor visibility, unreadable formats or irrelevant content were removed. The remaining images were standardized by resizing them to 224×224 pixels, which matches the input requirement of MobileNetV2. Pixel values were normalized using rescaling before being passed to the model.

Following preprocessing, the dataset exhibited class imbalance with a higher number of unsafe images. To mitigate this, the unsafe class was slightly down sampled. The resulting dataset contained 1,584 safe images and 1,799 unsafe images thus representing a near-balanced distribution.

Data augmentation techniques including random rotation, horizontal flipping, zooming and variations in the brightness of the image were applied to improve model generalization. The dataset was then split into training, validation and test sets using a hold-out validation strategy for model development and evaluation.

3.2 Model Architecture

A convolutional neural network based on transfer learning was employed for infant sleep posture classification. The MobileNetV2 architecture was selected as the backbone model due to its efficiency and suitability for edge deployment.

In the proposed architecture, each input infant sleep image is first resized to $224 \times 224 \times 3$ and normalized before being passed to the feature extraction stage. The extracted visual features are then converted into a compact representation using global average pooling which is followed by dropout regularization and final binary classification into safe and unsafe infant sleep postures through a sigmoid-activated dense layer.

The overall architecture of the proposed MobileNetV2-based model is illustrated in Fig. 1.

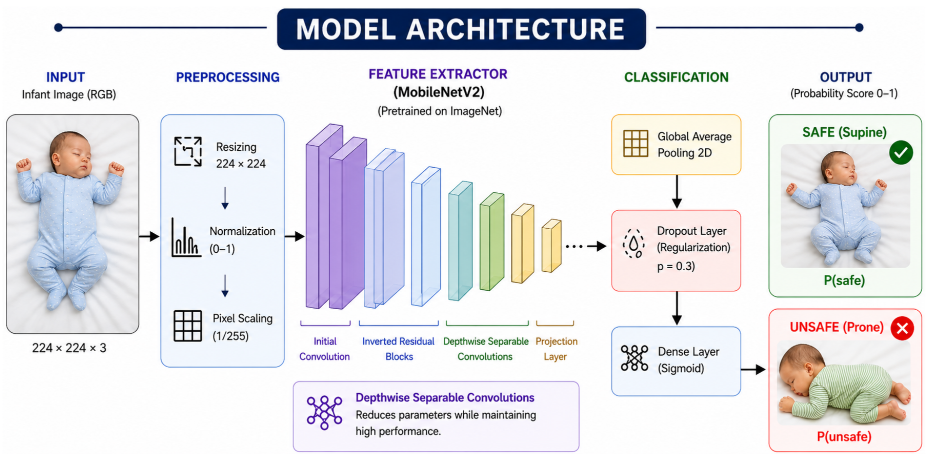

Fig. 1. Proposed MobileNetV2-based model architecture for infant sleep posture classification.

The pre-trained MobileNetV2 model, initialized with ImageNet weights, was used as a feature extractor. The base model layers were frozen to retain the learned representations while a custom classification head comprising of the global average pooling layer, dropout layer and sigmoid-activated dense layer was added for binary classification.

The MobileNetV2 backbone was frozen during training and only the final dense classification layer was optimized. This resulted in 1,281 trainable parameters out of 2,259,267 total parameters. The model was then trained using the binary cross-entropy loss function and optimized using the Adam optimizer.

The complete model architecture and training configuration are summarized in Table 1.

Table 1. Model architecture and training configuration

Component	**Details**
Framework	TensorFlow/Keras
Backbone	MobileNetV2 pretrained on ImageNet
Input Size	$224 \times 224 \times 3$
Preprocessing	Pixel rescaling by 1/255
Feature Map Output	$7 \times 7 \times 1280$
Pooling Layer	Global Average Pooling 2D
Regularization	Dropout ($p = 0.3$)
Classifier	Dense layer with 1 sigmoid unit
Total Parameters	2,259,267
Trainable Parameters	1,281
Non-trainable Parameters	2,257,984
Loss Function	Binary cross-entropy
Optimizer	Adam, default learning rate 0.001
Batch Size	32
Epochs	10

3.3 System Pipeline

The proposed system follows an end-to-end pipeline for infant sleep monitoring, integrating data preprocessing, model inference and deployment-oriented evaluation. It is designed to operate under both controlled and real-world conditions, enabling analysis of model behavior across static image inputs and real-time video streams including deployment on an edge device. The overall system pipeline consists of three stages: preprocessing, model inference and output generation and is illustrated in Fig. 2.

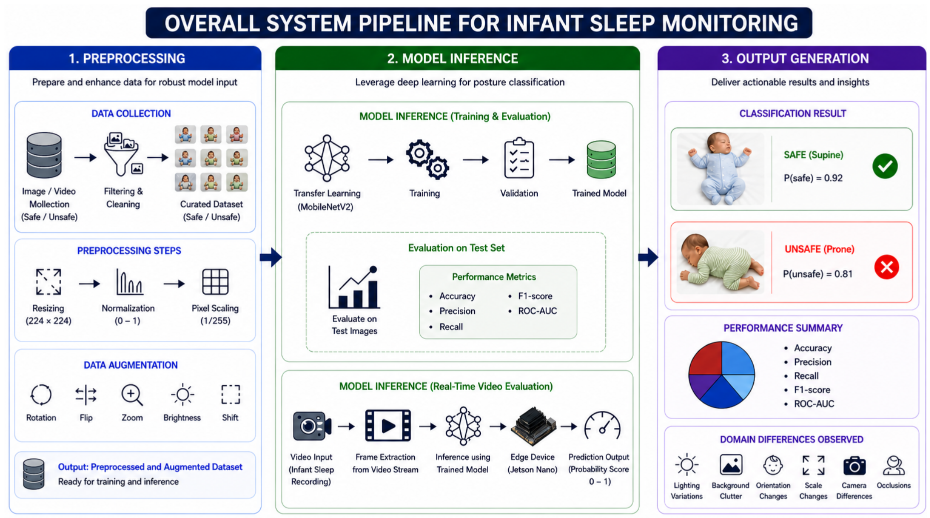

Fig. 2. Overview of the proposed system, including dataset preparation, model training, testing and real-time deployment on an edge device.

During preprocessing, input images or video frames are resized and normalized and then passed to the model. For real-time evaluation, the frames are extracted from the video input and processed individually.

The trained model then determines whether the posture is safe or unsafe. The output is a probability score. A score greater than or equal to 0.5 indicates an unsafe sleeping position while a score less than 0.5 indicates a safe sleeping position.

The predictions are overlaid on video frames along with confidence scores and latency measurements. Additionally, logs are generated to record frame-level predictions and inference times, thus enabling analysis of the model's behavior when it receives a continuous stream of input.

In addition to software-based evaluation, the system was examined through preliminary hardware implementation on an NVIDIA Jetson Nano edge device. In this setup, camera input was processed in real time and passed through the trained model for infant sleep posture classification. The proposed hardware design also included peripheral components such as an OLED display and a buzzer for system monitoring and alert generation, as illustrated in Fig. 3. However, the present evaluation primarily focused on camera-based inference and model behavior on the edge device.

To further illustrate the hardware integration, the circuit-level setup used for edge deployment is shown in Fig. 3.

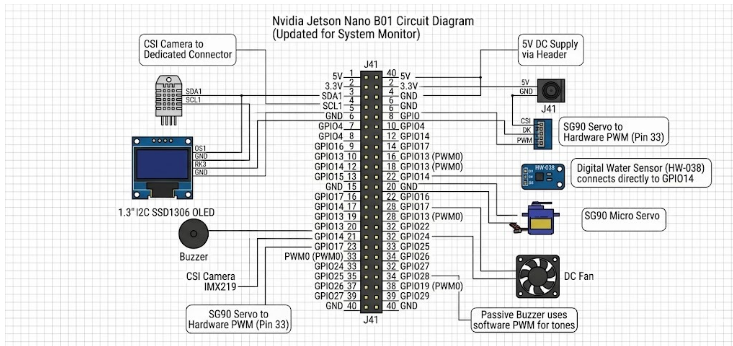

Fig. 3. Circuit diagram with the hardware setup used for edge deployment of the proposed system.

4 Results and Discussion

This section presents the results from the model and explains how it performs in both controlled and real-world situations. Along with the test dataset results, observations from video-based testing are also included.

4.1 Model Performance

The model was tested on a separate dataset to evaluate how well it can classify infant sleep postures as safe or unsafe. On the image dataset, it performed well, which shows that it can learn useful visual features related to posture under controlled conditions.

The confusion matrix obtained from the test set is summarized below and visualized in Fig. 4:

- True Safe (TN): 485
- False Unsafe (FP): 11
- False Safe (FN): 10
- True Unsafe (TP): 526

These values are calculated across the full test dataset. Based on this, the performance metrics are shown in Table 2. The high accuracy and ROC-AUC values show that the model can clearly distinguish between the two classes. The number of incorrect predictions is low, thus showing a balanced performance in controlled environments. This includes a low number of false positive and false negative predictions.

From these results, the performance metrics are summarized in Table 2.

Table 2. Performance results obtained from the test dataset.

Metric	Value
Accuracy	0.9797
Precision (Unsafe)	0.9795
Recall (Unsafe)	0.9813
F1-score (Unsafe)	0.9804
ROC-AUC	0.9964

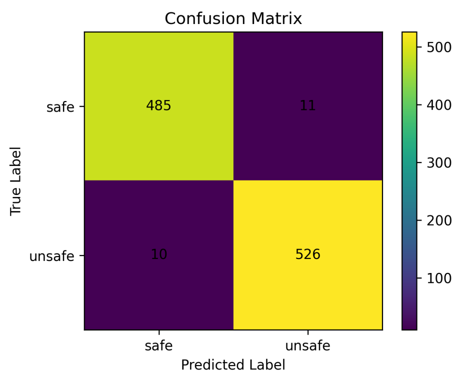

Fig. 4. Confusion matrix obtained from predictions made on the test dataset.

4.2 Real-Time Evaluation

To evaluate how the model performs outside of a controlled setup, testing was done using video input. Frames were extracted from the video and passed to the model one by one. The predictions along with confidence scores and processing time were observed for each frame.

In addition to software-based evaluation, the system was examined through preliminary hardware implementation on an edge device (NVIDIA Jetson Nano). In this setup, camera input was processed in real time and passed through the trained model for inference.

While the model demonstrated consistent performance on static image data as shown in Fig. 5, observations from the video-based evaluation indicated variations in prediction behavior. In particular, certain frames from the proxy evaluation setup resulted in predictions that differed from expected outcomes despite visually consistent posture cues. In such cases, the model classified inputs as unsafe with moderate confidence despite the posture appearing visually consistent with safe conditions. This behavior highlights the influence of domain differences between the training data and deployment environment. An example of such a case was observed during the evaluation.

Fig. 5. Examples of correctly classified safe (left) and unsafe (right) infant sleep postures with corresponding model confidence scores.

4.3 Discussion

The differences seen during video testing can be explained by the gap between the training data and the actual testing environment. For ethical, medical and safety reasons, the evaluation of the deployed system was performed using a doll rather than a real infant, thereby avoiding any unnecessary risk while enabling preliminary assessment of the system.

The usage of a doll instead of a real infant during testing changed visual features such as texture and shape which could not be fully represented in the training data which consisted of images of real infants. Also, camera angle, distance and framing clearly affect how posture-related features are captured.

Due to these factors, the model exhibited variations in the predictions made when the input is different from what it has seen during training. This suggests that strong performance on curated image datasets does not always translate to comparable performance under real-world deployment conditions.

In order to examine this further, an experiment was performed within a simulated setup. The experiment is described as follows:

- The probabilities of the video frames immediately preceding and succeeding the wrongly classified frame are taken into account and considered indicative of the true state in this simulation. The recorded video sequence showcases a doll placed in a box in order to simulate the sleeping position of a baby in a crib.
- Given the threshold where values greater than 0.5 are regarded as unsafe, it can be said that there will be a spike in the probability value of the frame where the erroneous prediction is made.
- If that single frame alone crosses the threshold for being unsafe while its neighbors are within the threshold for the safe category, the probability spike can be suppressed as an anomaly and corrected to match the average of the previous and succeeding frame's probability values.
- Thus, the erroneous frame is rectified and the model can make the correct prediction, leading to a reduction in the rate of false positives i.e. the identification of safe sleeping positions as unsafe positions.

This simulation is based on the idea of temporal continuity in physical systems, where physical states generally change gradually over a period of time rather than instantaneously changing between consecutive observations.

While this experiment was exploratory in nature and was not implemented as a fully trained Physics-Informed Neural Network (PINN), the observations suggest that incorporating temporal consistency constraints inspired by PINN-based approaches may help reduce such isolated false positive predictions. This is based on the observation that an isolated prediction spike occurring between neighbouring frames with otherwise consistent predictions is more likely to represent a transient misclassification than a genuine posture transition. Therefore, temporal consistency between adjacent frames provides the basis for the exploratory correction investigated in this work.

Fig. 6 shows an example of a false positive prediction and the corresponding conceptual correction.

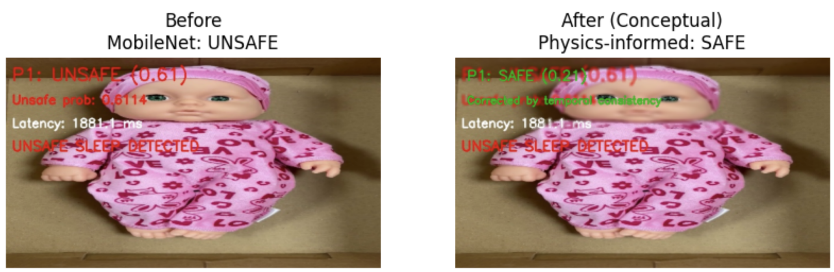

Fig. 6. Example of a false positive prediction before and after correction.

The correction process is also illustrated as a plot in Fig. 7.

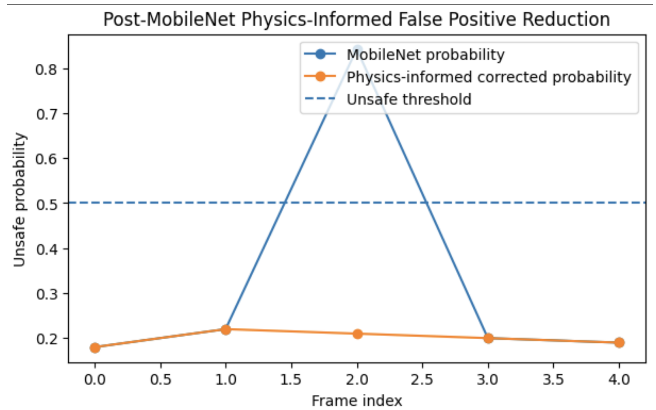

Fig. 7. Illustration of suppression of a false positive using temporal consistency-based correction.

5 Conclusion and Future Scope

This work explored a deep learning-based approach for infant sleep monitoring with a focus on SIDS risk assessment. The model performed well on the dataset, achieving an accuracy of 97.97% and a ROC-AUC of 0.9964, demonstrating its ability to classify infant sleep postures correctly under controlled experimental conditions.

However, during real-time testing, the performance differed due to variations in lighting, background and overall visual conditions which had a noticeable impact on the predictions.

An exploratory temporal consistency-based correction experiment indicated that incorporating continuity constraints may help reduce isolated false positives i.e. cases where safe sleeping positions are incorrectly classified as unsafe. These findings suggest that improving real-world performance will require more diverse training data and better handling of environmental variations.

Future work will focus on using real-world video data, improving how the model handles different environmental conditions and further exploring sequence-aware and Physics-Informed Neural Network (PINN)-based approaches. In addition to this, more extensive experimentation and carefully monitored real-world trials conducted in compliance with ethical and clinical standards will be essential for enabling the safe deployment of such systems in practical healthcare settings.

References

1. Moon, R.Y., Carlin, R.F., Hand, I.: Evidence base for safe infant sleeping environment to reduce SIDS risk. Pediatrics **150**(1), e2022057991 (2022)

2. Centers for Disease Control and Prevention (CDC): Sudden unexpected infant death and sudden infant death syndrome. https://www.cdc.gov/sudden-infant-death/about/index.html, last accessed 2026/06/26
3. Greco, L., Percannella, G., Ritrovato, P., Tortorella, F., Vento, M.: Trends in IoT-based solutions for healthcare: moving AI to the edge. Pattern Recognition Letters **135**, 346–353 (2020)
4. Ananta, E.D., Syaifuddin, Soetjiatie, L., Utomo, B.: Development of IoT-based infant monitoring system for preventing SIDS. Jurnal Teknokes **16**(2), 73–79 (2023)
5. Jabbar, W.A., Shang, H.K., Hamid, S.N.I.S., et al.: IoT-based baby monitoring system for smart cradle. IEEE Access **8**, 93791–93805 (2020)
6. Stern, L., et al.: In-bed posture classification using deep neural networks. Sensors **23**(5), 2430 (2023)
7. Beno, L., Kucera, E., Basista, M.: Smart sleep monitoring: An integrated application for tracking and analyzing babies' sleep—BabyCare. Electronics **13**(21), 4210 (2024). https://doi.org/10.3390/electronics13214210
8. Jayakumar, S., Ranjith Kumar, R., Tejswini, R., Kavil, S.: IoT-based health monitoring system. Advances in Parallel Computing **39**, 193–200 (2021)
9. Ramadhani, U.A., Wisana, I.D.G.H., Nugraha, P.C.: Smartphone-based respiratory signal monitoring and apnea detection. Jurnal Teknokes **14**(2), 49–55 (2021)
10. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: Daumé III, H., Singh, A. (eds.) Proceedings of the 37th International Conference on Machine Learning (ICML), Proceedings of Machine Learning Research (PMLR), vol. 119, pp. 1597–1607 (2020)