Title: Transitioning sleeping position detection in late pregnancy using computer vision from controlled to real-world settings: an observational study

Authors: Allan J Kember, Hafsa Zia, Praniya Elangainesan, Min-En Hsieh, Ramak Adijeh, Ivan Li, Leah Ritchie, Sina Akbarian, Babak Taati, Sebastian R Hobson, Elham Dolatabadi

Supplementary File 1: Additional Methodological Details

Assignment of Validation Set and Testing Set for Each Loop of the Cross Validation

For each loop, i, of the five-fold cross validation, the i^{th} fold was assigned as the testing set, and the $(i+1)^{th}$ fold was assigned as the validation set, while the remaining folds served as the training set, resulting in the sequence of validation, testing, and training sets shown in **Table S1.1**.

Allocation of Participants to Five Folds

For the SLeeP AIDePt-2 k-fold cross-validation, we set k=5 (i.e., we split the data up into five non-overlapping folds), which was different from SLeeP AIDePt-1 where we set k=6.[1] Despite this difference, we used five (folds 1, 2, 3, 4, and 5) of the same non-overlapping folds as SLeeP AIDePt-1, and distributed the four participants from the sixth fold (fold 6) of SLeeP AIDePt-1 across four folds (folds 1, 3, 4, and 5) of SLeeP AIDePt-2. Note that each participant's entire dataset (i.e., the annotated frames containing the participant) was completely allocated to only one fold, that is, their dataset was not divided up across folds, ensuring that no two folds contained data from the same participant. As such, folds 1, 3, 4, and 5 of SLeeP AIDePt-2 each contained the annotated frames of five participants from the simulated position dataset, and fold 2 contained the annotated frames of four participants from the simulated position dataset (see **Table S1.2**). We then added the annotated frames from at least two participants with bed partners from the real-world dataset to each of the five folds, and distributed the remaining real-world single-participant and multi-participant datasets across the five folds. In sum, this resulted in each of the five non-overlapping folds containing frames of sleeping positions from approximately five unique single-participant sleeping settings and three unique multi-participant sleeping settings. Note that we did not stratify the selection of the two real-world participants and bed partners based on the frequency of classes; however, we did try to ensure that each fold contained at least one real-world data participant and bed partner containing a household pet and pregnancy pillow.

Tables

Table S1.1. Assignments of validation set, testing set, and training set for each loop of the cross validation

Loop	Testing set	Validation set	Training set
1	Fold 1	Fold 2	Folds 3, 4, 5
2	Fold 2	Fold 3	Folds 4, 5, 1
3	Fold 3	Fold 4	Folds, 5, 1, 2
4	Fold 4	Fold 5	Folds 1, 2, 3
5	Fold 5	Fold 1	Folds 2, 3, 4

Table S1.2. Number and origin (simulated position dataset vs. real-world dataset) of unique sleep settings (single-participant and multi-participant) in each fold of the five-fold cross validation

Fold	Single-participant sleep settings		Multi-participant sleep settings[a]
	Simulated position dataset	Real-world dataset	
1	5	0	3
2	4	1	3
3	5	0	2
4	5	0	3
5	5	1	3
Total	24	2	15

[a] Note that all the participants in multi-participant sleep settings were from the real-world dataset.

References

1. Kember, A. J. *et al.* Vision-based detection and quantification of maternal sleeping position in the third trimester of pregnancy in the home setting–Building the dataset and model. *PLOS Digit. Health* **2**, e0000353 (2023).

Supplementary File 2: Additional Results

Loop-wise Summary of SLeeP AIDePt-2 Model Training and Validation

A summary of the training and validation of SLeeP AIDePt-2 is given in **Table S2.1**. The summary is "loop-wise" in that it includes, for each loop of the cross validation, the number of validation images and instances (i.e., number of annotations contained in the frames in the validation set), the number training epochs completed prior to activation of the early stopping criteria, the epoch with the best results (highest validation mAP@0.50), and the performance parameters (validation precision, validation recall, validation mAP@0.50, and validation mAP@.50-.95) corresponding to the best results and across all classes.

Loop-wise Summary of SLeeP AIDePt-2 Model Performance Testing

The best weights from each loop of the cross validation (i.e., the weights that achieved the highest validation mAP@0.50 during training and validation) were used to evaluate performance on the test set for each loop. A summary of the performance results of each of the five models in SLeeP AIDePt-2 produced from the cross-validation is given in **Table S2.2**. The summary is "loop-wise" in that it includes, for each loop of the cross validation, the number of frames in the test set, number of test instances (i.e., number of annotations contained in the frames in the test set), and performance parameters (precision on test set, recall on test set, mAP@0.50 on test set, and mAP@.50-.95 on test set) across all classes.

Tables

Table S2.1. Summary of SLeeP AIDePt-2 Training and Validation of Five Models via Five-fold Cross Validation with Performance Averaged Across All Classes

Loop of CV	No. of images in val set	No. of instances in val set	Number of epochs completed	Best results at epoch No.	Precision on val set	Recall on val set	mAP@0.50 on val set	mAP@.50-.95 on val set
1	2245	9195	M	M	0.765	0.554	0.665	0.533
2	2619	9742	M	M	0.733	0.690	0.717	0.567
3	3075	10963	M	M	0.705	0.557	0.580	0.447
4	2251	7152	M	M	0.673	0.614	0.656	0.496
5	2740	11317	64	44	0.617	0.577	0.615	0.475

Abbreviations: CV indicates cross validation; M indicates missing value due to disconnection of the Colab GPU during model training and validation; mAP@0.50 indicates the mean average precision at an intersection over union threshold of 0.50; mAP@.50-.95 indicates the mean average precision averaged across multiple intersection over union thresholds between 0.5 and 0.95; No. indicates number; val indicates validation.

Table S2.2. Summary of SLeeP AIDePt-2 Testing of Five Models via Five-fold Cross Validation with Performance Averaged Across All Classes[a]

Loop of CV	No. frames in test set	No. of instances in test set	Precision on test set	Recall on test set	mAP@0.50 on test set	mAP@.50-.95 on test set
1	2740	10963	0.647±0.168	0.719±0.175	0.706±0.178	0.550±0.137
2	2245	8695	0.722±0.148	0.653±0.168	0.724±0.144	0.586±0.114
3	2594	8884	0.692±0.197	0.706±0.186	0.710±0.186	0.542±0.133
4	3075	10871	0.705±0.154	0.565±0.270	0.633±0.217	0.483±0.159
5	2233	7091	0.794±0.118	0.643±0.181	0.712±0.198	0.540±0.152

Abbreviations: CV indicates cross validation; mAP@0.50 indicates the mean average precision at an intersection over union threshold of 0.50; mAP@.50-.95 indicates the mean average precision averaged across multiple intersection over union thresholds between 0.5 and 0.95; No. indicates number.

[a] Performance parameters displayed as mean ± standard deviation because they are averaged across all classes.

Supplementary File 3: Additional Comments

Limitations

Camera Placement

It is important to note that SLeeP AIDePt-2 is trained on frames extracted from video recordings achieved by attaching the camera to the wall at the head of the bed, centred, and 1.6-1.7 metres above the sleeping surface; however, this placement is not possible in some circumstances (e.g., low ceiling height or wall shelving). When this placement is not achieved, the camera orientation and perspective may differ from the predominant orientation and perspective in our underlying training dataset, which is an orientation such that participants' bodies are oriented vertically in the image with their feet at the top and head at the bottom and a perspective where the camera is looking down from almost directly above the participants' heads and centred between them. See **Figure S3.1** for an example of the image orientation and perspective from correct and incorrect placement of the camera.

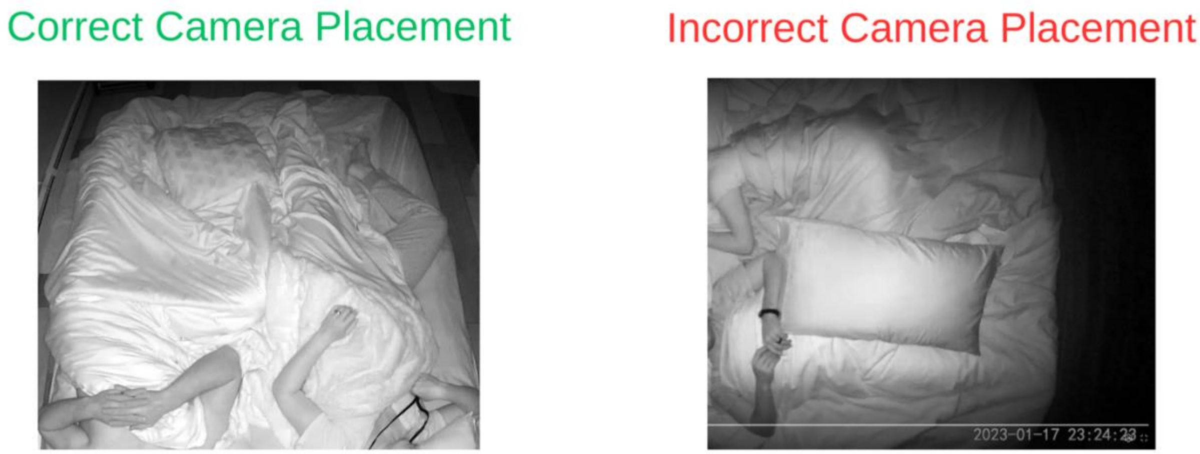

Figure S3.1. Impact of camera placement on image orientation and perspective. The correct image orientation and perspective from correct camera placement is shown on the left panel. Incorrect image orientation and perspective from incorrect camera placement is shown on the right panel.

As noted in the **Results** section of the main text, two participants (and their bed partners) installed the camera in the wrong orientation, so their data were excluded from model building. However, we performed a “challenge test” on SLeeP AIDePt-2 by testing its performance on a dataset from one of these participants (and bed partners) who inadvertently placed the camera incorrectly (“challenge dataset”) in an effort to assess how model performance is impacted by camera placement. For the challenge dataset, the camera was mounted on the wall beside the bed rather than on the wall at the head of the bed, which gave a sideways image orientation (90 degrees clockwise rotation such that the participants’ heads were at the left of the image) and perspective where the camera was looking down from almost directly above the bed partner’s body (see right panel in **Figure S3.1** for an example frame from the challenge dataset).

The challenge dataset contained a total of 360 frames and 1,614 annotations. When this dataset, as is, was passed into SLeeP AIDePt-2, the model performance was dire and its predictions were no better than chance alone. Therefore, we decided to rotate the frames and annotations by 90 degrees counterclockwise in order to get them in the correct orientation prior to passing them into SLeeP AIDePt-2. With this preprocessing operation, SLeeP AIDePt-2 achieved the results shown in **Table S3.1**.

From these results, it is evident that the performance of SLeeP AIDePt-2 is sensitive to both orientation and perspective of the frames. While we corrected the orientation of the challenge dataset frames by rotating them, we were unable to correct the perspective. As such, the performance of SLeeP AIDePt-2 on the challenge dataset was significantly lower than we observed during formal model testing, achieving AP@0.50's for the left lateral, supine, and right lateral classes of 0.74 (vs. 0.89), 0.63 (vs. 0.82), and 0.65 (vs. 0.84), respectively. The false negative rate for these three classes was quite high, ranging from 29 to 43% (recall values 0.57 to 0.71). Other classes (besides pillow) were poorly detected. Note that a zero value for recall (sensitivity) for a given class indicates a 100% false negative rate (the model did not detect the class when it was present). A zero value for precision for a given class means that every time the model detected that class it was incorrect, whereas a value of 1.00 indicates that every time the model detected that class its detection was correct.

In sum, while performance of SLeeP AIDePt-2 on the challenge dataset could be impacted negatively by other factors (e.g., the challenges inherent in the data containing this particular couple, or the difference in size [number of pixels] of the peoples' bodies and bed relative to the size of the frame from filming in a 90 degree rotated orientation and filling more of the frame), we believe that one of the main contributing factors is the perspective of the frames from incorrect camera placement.

Exclusion of Accelerometry Data From Model Testing Proper

During the performance evaluation of SLeeP AIDePt-2, it is crucial to note that we compared the predictions made by each model on the test set with the ground truth annotations for that test set. Specifically, we did not compare the SLeeP AIDePt-2 predictions against the accelerometry data (NightOwl sensors attached to the participant's and bed partner's abdomen) for several reasons. First, the sensor data collection failed on some nights, so the dataset was not complete. Second, the sensor was donned upside down by the participant or bed partner on some nights, which reversed the left and right measurements. Third, such a comparison would be an "apples-to-oranges" comparison because of the low resolution of the sensor (it only detects left, right, prone, supine, and upright postures) in contrast to our model (twelve sleeping positions and sitting); as such, the sensor may not account for pelvis position since it was placed near the xiphoid process on the upper abdomen and the spine can be twisted such that the pelvis position and upper abdomen position are incongruent. Furthermore, we are not privy to the sensor manufacturer's angle (degrees) cutoff between a lateral position (e.g., right) and a non-lateral position (e.g., supine) and do not yet have access to the raw accelerometry data.

See **Supplementary File 1** the 'Dataset Development' section in the Methods in the main text for more details regarding the specific circumstance for our use of the accelerometry data during model building.

Controversy and Uncertainty Regarding an Association Between Maternal Supine Sleeping Position and Adverse Pregnancy Outcomes

There is some controversy and uncertainty in the literature regarding the association between maternal supine sleeping position, stillbirth, and giving birth to a small-for-gestation (SGA) infant.[1–5]

Regarding uncertainty, in an evidence review by the National Institute of Healthcare and Excellence (NICE) and the Royal College of Obstetricians and Gynaecologists (RCOG), the review concluded, "*While the quality of the evidence from the primary studies ranged depending on the timing and precise outcomes considered, the quality of the evidence from the [individual patient data] IPD, particularly for the evidence around supine sleeping position, was relatively high… However the committee agreed it was of sufficient quality to advise women to try to avoid going to sleep on their back after 28 weeks and inform women of the likely link with stillbirth, alongside a caveat that the evidence is uncertain. The committee chose to specifically highlight stillbirth as this is a more concerning outcome than babies being born SGA and they agreed that including SGA in the recommendations made the advice less clear.*"[5] As a result of this NICE/RCOG review, the RCOG Antenatal Care Guideline does not mention an association between maternal supine sleep and giving birth to an SGA infant.[6]

Regarding controversy, this stems from a study by Silver et al.[1] and the commentary it has generated.[1–4] The study by Silver et al. is one of six studies in the literature investigating the association between supine sleep and adverse pregnancy outcomes.[1,7–11] Of these six studies, Silver et al. is the only prospective study (whereas the other five are retrospective). Of these six studies, Silver et al. is the only study to have found no association between supine sleep and adverse pregnancy outcomes. The source of the controversy is not the disparate findings of Silver et al. but rather the reaction of the some in the medical community (Fox and Oster) to have weighed the evidence of the Silver et al. study above the other five studies (because it was prospective and they were retrospective) and dismissed the association between supine sleep and adverse pregnancy outcomes reported in the other five studies.[4] However, Fox and Oster failed to notice a major difference in Silver et al.'s methodology:[2,3] whereas the other five studies looked at sleeping position from 28 weeks' gestation through birth, Silver et al. looked at sleeping position from 20 to 30 weeks' gestation, which makes for an apple-to-oranges comparison to the other five studies and makes it difficult to draw definitive conclusions about an association between sleeping position from 28 weeks' to birth and adverse pregnancy outcomes.

Future Research: the DOSAGE Study

The DOSAGE (Dose Of Supine sleep Affects fetal Growth? an Exposure-response) study will be an international, prospective, cohort study aiming at gathering objective evidence that will either lend support to (or detract support from) a causal link between supine sleeping position after 28 weeks' gestation, foetal growth, and late stillbirth, and to quantify the safe "dose" of nightly supine sleeping time, if any.

Eligibility criteria includes ≥28 weeks' gestation, sleeps in a bed a night, possesses or can easily procure a home-security camera, and bed partner, if any, consents to participate.

Recruitment will stop when either of the following are achieved: (1) Maximum N=986 participants reached (two-sample t-test, Cohen's D 0.2, significance 0.05, power 0.8, with 20% dropout), or (2) Stopping criteria are met per Bayesian methodology (Bayes factor obtained; $BF_{10} \geq 10$ or $BF_{01} \leq 10$).[12]

Participants will record their sleep one night per week from 28 weeks to birth using any commercially-available home-security camera mounted above the head of their bed. Participants will upload their videos to a secure, online portal where SLeeP AIDePt-2 will automatically detect and quantify the amount of time spent in each sleeping position, which, averaged over approximately twelve weeks, will act as a surrogate of the average doses of time spent in each sleeping positions per night across the third trimester. After birth, participants will self-report pregnancy, labour, and birth outcomes. See the following YouTube link for a video of the data collection and transfer process: https://youtu.be/Vjh2wgegYEs

The primary outcome, customised birthweight centile per the Gestation-Related Optimal Weight standard (https://www.gestation.net/cc/about.htm), will be computed and regressed on sleep position dosages.

Tables

Table S3.1. Performance Results of Challenge Test of SLeeP AIDePt-2 on the Challenge Dataset[a]

Class	No. of instances[b]	Precision	Recall	AP@0.50	AP@.50-.95
All[c]	1614	0.485±0.326	0.288±0.276	0.282±0.272	0.159±0.166
Left recovery	46	0.423±0.075	0.200±0.079	0.219±0.025	0.108±0.018
Left lateral	236	0.700±0.035	0.698±0.189	0.735±0.112	0.456±0.103
Left tilt	5	0.000±0.000	0.000±0.000	0.011±0.009	0.003±0.004
Supine	123	0.641±0.127	0.572±0.142	0.628±0.025	0.321±0.021
STwLPT	22	0.205±0.177	0.058±0.072	0.082±0.059	0.036±0.021
STwRPT	10	0.189±0.020	0.200±0.100	0.129±0.061	0.064±0.039
SPwLTT	12	1.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000
SPwRTT	3	0.800±0.447	0.000±0.000	0.000±0.000	0.000±0.000
Right tilt	3	0.097±0.092	0.200±0.182	0.156±0.175	0.090±0.107
Right lateral	141	0.595±0.125	0.713±0.054	0.649±0.079	0.343±0.088
Right recovery	53	0.460±0.112	0.429±0.052	0.370±0.049	0.222±0.027
Prone	27	0.577±0.393	0.200±0.228	0.181±0.115	0.081±0.040
Sitting	20	0.436±0.264	0.140±0.022	0.134±0.049	0.062±0.023
Pillow	913	0.661±0.123	0.624±0.032	0.660±0.040	0.437±0.053

Abbreviations: AP@0.50 indicates the average precision at an intersection of union of 0.50; AP@.50-.95 indicates the average precision at intersections of unions between 0.50 and 0.95; No. indicates number; SPwLTT indicates supine pelvis with left thorax tilt; SPwRTT indicates supine pelvis with right thorax tilt; STwLPT indicates supine thorax with left pelvic tilt; STwRPT indicates supine thorax with right pelvic tilt.

[a] Performance metrics results (precision, recall, AP@0.50, AP@0.50-0.95) are averages of the results from the best model from each loop of the five-fold cross validation of SLeeP AIDePt-2 and are presented as mean ± standard deviation.

[b] The number of instances is the number of times the class is annotated in the dataset.

[c] The value in the AP@0.50 column is the mean AP@0.50, and the value in the AP@.50-.95 column is the mean AP@.50-.95 since this row represents averages across all classes.

References

1. Silver, R. M. *et al.* Prospective Evaluation of Maternal Sleep Position Through 30 Weeks of Gestation and Adverse Pregnancy Outcomes. *Obstet. Gynecol.* **134**, 667–676 (2019).

2. McCowan, L. M. E. *et al.* Prospective Evaluation of Maternal Sleep Position Through

30 Weeks of Gestation and Adverse Pregnancy Outcomes. *Obstet. Gynecol.* **135**, 218 (2020).

3. Silver, R. M., Reddy, U. M. & Gibbins, K. J. Response to Letter. *Obstet. Gynecol.* **135**, 218–219 (2020).
4. Fox, N. S. & Oster, E. F. The Advice We Give to Pregnant Women: Sleep On It. *Obstet. Gynecol.* **134**, 665 (2019).
5. National Institute for Health and Care Excellence, National Guideline Alliance & Royal College of Obstetricians and Gynaecologists. *Antenatal Care: [W] Maternal Sleep Position during Pregnancy - NICE Guideline NG201 - Evidence Reviews Underpinning Recommendations 1.3.24 to 1.3.25*. (2021).
6. Overview | Antenatal care | Guidance | NICE. https://www.nice.org.uk/guidance/ng201.
7. Stacey, T. *et al.* Association between maternal sleep practices and risk of late stillbirth: a case-control study. *BMJ* **342**, d3403 (2011).
8. Owusu, J. T. *et al.* Association of maternal sleep practices with pre-eclampsia, low birth weight, and stillbirth among Ghanaian women. *Int. J. Gynaecol. Obstet. Off. Organ Int. Fed. Gynaecol. Obstet.* **121**, 261–265 (2013).
9. McCowan, L. M. E. *et al.* Going to sleep in the supine position is a modifiable risk factor for late pregnancy stillbirth; Findings from the New Zealand multicentre stillbirth case-control study. *PLoS ONE* **12**, e0179396 (2017).
10. Heazell, A. *et al.* Association between maternal sleep practices and late stillbirth - findings from a stillbirth case-control study. *BJOG Int. J. Obstet. Gynaecol.* **125**, 254–262 (2018).

11. O'Brien, L. M. *et al.* Maternal sleep practices and stillbirth: Findings from an international case-control study. *Birth Berkeley Calif* **46**, 344–354 (2019).

12. van Doorn, J. *et al.* The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychon. Bull. Rev.* **28**, 813–826 (2021).