www.nature.com/pr

CLINICAL RESEARCH ARTICLE OPEN

Automated neonatal sleep positioning assessment by video monitoring and machine learning

Péter Földesy[1]✉, Judit Varga[2], Ádám Nagy[1], Flóra Fehér[1], Zita Lilla Róka[1], Ákos Antal[1,3], Imre Gergely Jánoki[1,3] and Miklós Szabó[2]

BACKGROUND: Proper positioning during sleep is critical for musculoskeletal and neurological development of preterm neonates, yet current manual assessment methods are time-intensive and subject to inter-observer variability.
METHODS: In this proof-of-concept study, we retrospectively investigated body alignment (supine, prone, right lateral, left lateral, semi-prone) and positioning of infants during sleep episodes. We used the Infant Positioning Assessment Tool (IPAT) scoring from recordings obtained during daily routine care at two NICUs. Anatomical keypoints were annotated manually and used for fine-tuning the ViTPose pose estimator and as input for machine learning classifiers to predict alignment and suboptimal positioning.
RESULTS: A total of 1510 images from 50 stable preterm neonates during their quiet periods in closed incubators (gestational age 23–36 weeks) were assessed. Posture classification of five categories achieved a macro F1-score = 0.80 (95% CI: 0.74, 0.86). Suboptimal positioning prediction reached AUROC = 0.86 (0.79, 0.87) and AUPRC = 0.69 (0.59, 0.79). Input image to decision arrangement yielded AUROC = 0.87 (0.84, 0.92) and AUPRC = 0.70 (0.61, 0.82).
CONCLUSIONS: Video-based automated posture and positioning assessment is feasible, enabling continuous monitoring and early warning of positioning-related complications.

Pediatric Research; https://doi.org/10.1038/s41390-026-05237-4

IMPACT:

- Positioning of preterm neonates is critical for musculoskeletal and neurological development; however, no automated visual assessment model is available.
- Visual machine learning models can assess infant positioning and predict repositioning needs.
- Performance saturates across architecturally diverse classifiers, indicating that further gains require improved labeling consistency rather than more complex models.
- Continuous 24/7 NICU incubator camera images can be used to detect positioning deterioration between routine nursing assessments.

INTRODUCTION

Therapeutic positioning of preterm infants in the Neonatal Intensive Care Unit (NICU) is critical for musculoskeletal and neurological development. Without developmentally supportive positioning, neonates face risks including skeletal deformities, respiratory distress, motor disabilities, and cognitive impairment, with these complications being particularly severe in extremely preterm infants.[1–4] The appropriate positioning of premature infants, often referred to as Neonatal Supportive Positioning, is crucial for promoting neurosensory and emotional development and mitigating stress in the NICU.[5,6] To standardize best practices for positioning and objectively evaluate posture, several specialized assessment tools have been developed for this population. However, current clinical assessment relies on intermittent visual evaluation by trained staff, which introduces inter-rater variability and limits monitoring frequency.[7]

Neurodevelopmental support programs, such as the Newborn Individualised Developmental Care and Assessment Program[8] and Family and Infant Neurodevelopmental Education[9] incorporate positioning assessments as a core component of their comprehensive developmental care frameworks. The original Infant Positioning Assessment Tool (IPAT), introduced by Coughlin and colleagues in 2010[10], provides a validated visual scaling framework for evaluating body alignment across supine, prone, and lateral positions. Despite its variants and recognition as a robust objective measure,[10] IPAT manual implementation faces practical limitations. Assessment requires targeted educational programs and trained personnel to perform time-intensive visual evaluations at discrete intervals, introducing inter-rater variability and constraining monitoring frequency.[11,12] The assessment competes with other clinical priorities, potentially resulting in suboptimal monitoring and delayed detection of positioning issues.

[1]HUN-REN Institute for Computer Science and Control – SZTAKI, Budapest, Hungary. [2]Department of Neonatology, Pediatric Center, Semmelweis University, Budapest, Hungary. [3]Faculty of Information Technology and Bionics, Péter Pázmány Catholic University, Budapest, Hungary. ✉email: foldesy.peter@sztaki.hun-ren.hu

Received: 16 December 2025 Revised: 22 April 2026 Accepted: 9 June 2026
Published online: 26 June 2026

Recent advances in computer vision and machine learning have created opportunities for automated clinical monitoring. Several deep learning architectures have been developed for neonatal applications.[13–16] Infant pose-estimation algorithms detect anatomical keypoints and construct skeleton-like models; however, their application to systematic positioning assessment has not been evaluated.

Our objectives were twofold: (1) identify optimal machine learning approaches for infant position classification from skeletal features; (2) determine whether a solution providing direct image to suboptimal positioning indication is reachable with current techniques.

METHODS

Our choice for IPAT is based on its reliability with the clear possibility for automation. The IPAT evaluates six anatomical features: head and neck alignment, trunk alignment, arm positioning, leg positioning, overall symmetry, and leg anatomic position, each scored 0 (poor) to 2 (optimal), for a maximum of 12 points. Scores of 0–8 indicate suboptimal positioning requiring intervention, 9–11 represent acceptable positioning, and 12 indicates ideal positioning. In addition, five distinct body alignments are also classified (supine, prone, right lateral, left lateral, semi-lateral position (≈45° between supine and lateral)). A custom body model was developed, marked with 23 anatomical keypoints for each image, in order to be able to assess detailed body part relations.

A total of six annotators participated in image scoring and keypoint labeling, including both clinical professionals and trained non-professionals. Annotators underwent standardized training consisting of on-site structured instructional sessions, exercises using a reference subset, supervised practice rounds to ensure consistent application of scoring criteria, and inter-rater reliability. Final inter-rater agreement after rater training and evaluation on a 156-case cohort yielded a Fleiss' κ of 0.69 for suboptimal positioning and 0.79 for body alignment. Total training load was approximately 5–8 h per annotator over 4–5 sessions. Based on the observed annotation agreement, the distribution of position cases, and the average class imbalance ratio of ~1:4 across IPAT scores of the trial cohort, we estimated the minimum required sample for ≤5% variation in reliability metrics $n > 1340$.[10] The annotated cohort data were integrated into the dataset of the study.

Study design and setting

This retrospective observational study was conducted in two physically separated Level III/IV NICUs within the same health institution at Semmelweis University, Hungary, from June 2023 to February 2025. The recordings and the study were approved by the Semmelweis University Research Ethics Committee (SE-RKEB: 265/2022). Continuous audio/video recordings were performed using a proprietary camera system. The system is comprised of off-the-shelf camera and single board computer components in a custom enclosure designed for closed incubator use, featuring active infrared illumination for 24-h recording capability in variable lighting conditions, including complete darkness. The videos were recorded at 1280×720 pixels resolution, 25 frames per second. All procedures were conducted in accordance with the Declaration of Helsinki. Data handling complied with EU General Data Protection Regulation requirements.

Participant eligibility and consent

Preterm infants (<37 weeks' gestational age at birth) admitted to participating NICUs were eligible. Parents/legal guardians provided written informed consent for video recording and data analysis prior to camera installation. Exclusions were applied in case of need for mechanical ventilation, hemodynamic/vasopressor support, centrally acting medications (opioid analgesics, sedatives), or presence of active infection. Infants with gestational age ≥37 weeks at birth and those with insufficient clinical data for analysis were also excluded.

Selection procedures

A total of 8400 h of video was captured from 88 enrolled infants during their NICU stay. Recording occurred continuously during periods when infants were housed in closed incubators receiving routine clinical care. From the image pool, two complementary datasets were constructed:

1. Random Surveillance Dataset: still images systematically extracted at 2-h intervals from all available video recordings across enrolled infants
2. Longitudinal Tracking Dataset: four infants representing key gestational age categories (24, 27, 30, and 33 weeks GA at birth) were purposively selected for intensive 24-h observation periods with image extraction every 10 min. The selection was based on GA variability and the length of available continuous recording, with minimal exclusion.

Individual still images were systematically extracted from video recordings. Images were retained for analysis only if they met all of the following technical criteria:

- Infant in documented active, quiet, or intermediate sleep stage (not awake or crying) to assess prolonged positions
- ≥70% of the body keypoints are visible without obstruction
- Infant positioned in closed incubator (images during open incubator care or radiant warmer use were excluded)
- Adequate image quality, as determined by the automated quality score threshold of brightness
- No ongoing caregiving activities

Images were excluded for: technical recording failures, empty bed frames, poor lighting/resolution, or incomplete body visualization.

Adaptive sampling enhancement

To ensure adequate representation of the full range of IPAT cumulative scores, we employed adaptive resampling. Preliminary analysis of the Random Surveillance Dataset identified underrepresented clinical cases ($0 \leq$ IPAT cumulative score ≤ 7). For infants and time periods exhibiting these rare cases, supplemental high-frequency sampling (1–10-min intervals) was performed within ±30 min of the identified events.

Sought outcome and variables

Models were trained to predict IPAT score and posture using body keypoint coordinates as input data. The model outputs were evaluated for posture classification (5 categories), IPAT categorical scores (6 features × 3 scores), cumulative scores (0–12), and repositioning requirements (binary yes/no). All output types were one-hot encoded as categorical data. No additional patient demographic or clinical data were included in model training.

Body keypoints and feature generation

The body keypoint specification was derived from ref. [17], including fingers and toes similar to the AggPose model.[18] The resulting body model comprised 23 independent keypoints. The inclusion of toes is justified by the anatomical leg position assessment requirements of the IPAT scale, while fingers enable more precise hand positioning analysis, such as detecting arm positions like hands touching the face. The list of keypoints and examples of body models and positioning classes is illustrated in Fig. 1.

Body keypoint coordinates were normalized using the maximum side length of the bounding box of the complete body.[19] A feature vector was derived from the normalized coordinates as the scalar input for the classifier models. The feature vector comprised the normalized coordinates and the following derived scalars: body part lengths, distances between non-connected joints, joint angles, and angles between segments (such as the angle between the two thighs). Feature engineering preceded the evaluation process by first including a comprehensive set of features, then removing redundant and low-influence ones.

Prediction models

Four machine learning algorithms were selected to span distinct inductive biases to predict posture and positioning quality from keypoints: Random Forest, eXtreme Gradient Boost (XGBoost), Support Vector Machine (SVM), and CatBoost. These algorithms represent complementary approaches: Random Forest uses parallel ensemble learning with independent decision trees; XGBoost employs sequential boosting where each tree corrects predecessors' errors; SVMs optimize margin-based separation using hyperplanes in high-dimensional space; and tree-based methods effectively capture geometric relationships and decision thresholds in pose data. Comparing them tests whether performance on this task is method-sensitive or instead converges to a ceiling imposed by annotation noise.

Models were trained and evaluated using Python 3 with Google Compute Engine. During training and validation, leave-one-infant-out cross-validation was used. A cycle ran through all infants; in each cycle, the models were trained on the remaining images, and their prediction performance was evaluated on the selected infant's images. The results of each run were finally combined using macro averaging for any imbalanced multi-class classification.

Model interpretability

To understand the black-box decision-making process of the classifiers, we employed the SHapley Additive exPlanations (SHAP) method.[20] We generated Shapley beeswarm importance data to interpret the magnitude of each feature's contribution to model output, revealing the relative importance and effect patterns of skeletal keypoint features. To enhance interpretability, we projected the derived feature set back onto the original anatomical keypoints of the infant body model. We illustrated their spatial distribution on the body schema, thereby linking models to specific postural landmarks. The importance map also highlights the weight of any missing keypoints in the decision.

Performance analysis

Following the Clinical AI Research recommendations,[21] the models were evaluated using the area under the Receiver Operating Characteristic curve (AUROC), area under the Precision-Recall curve (AUPRC), Fleiss' κ, precision, recall, F1-score, and confusion matrix.

During the classification of the repositioning required decision, we observed performance differences across the two prediction strategies: (a) separately predicting fine-grained IPAT component scores and comparing their cumulative sum to the threshold afterwards, (b) directly predicting repositioning indication as a binary outcome. To investigate the impact of these architectural choices on model performance, we present results for both cases.

Body model

We explored the usability of broadly used models that rely on different keypoint definitions as input for predictions. The first model we evaluated was the infant pose estimator AGMA-PESS, reported as the best-performing model in a recent evaluation.[22] The model was originally trained for pose estimation in controlled laboratory settings with standard human motion, rather than the challenging incubator environment featuring neonatal proportions, variable infrared lighting, medical equipment occlusions, and closed incubator reflections. The second model evaluated was ViTPose (ViTPose-huge variant),[23] a vision transformer-based architecture pre-trained on general human pose datasets. This model is the current state-of-the-art generic model on the COCO dataset and also the best-performing model for infant body modeling.[14,22] The model covers fewer keypoints than the assessment requires, necessitating extension. We successfully fine-tuned the pre-trained ViTPose model on our dataset with a new output generating head of 23 keypoints, achieving an average Keypoint Similarity of 0.73. To align the ViTPose model to our richer keypoint set, we integrated these additional keypoints into the model head and fine-tuned, which extracts keypoints from input images pre-processed by the backbone. The added coordinates are finger and toe keypoints (corresponding to points 9, 14, 18, and 22 in Fig. 1), the upper chest and navel keypoints (points 10, 23) to complement the shoulder and hip points, and finally, the head-top and neck points were included. The evaluation process is illustrated in Fig. 2.

From image to decision, called an end-to-end solution, is where this ViTPose model produces keypoints directly from raw incubator images; these keypoints feed the CatBoost classifier for training, which outputs the suboptimal positioning indication.

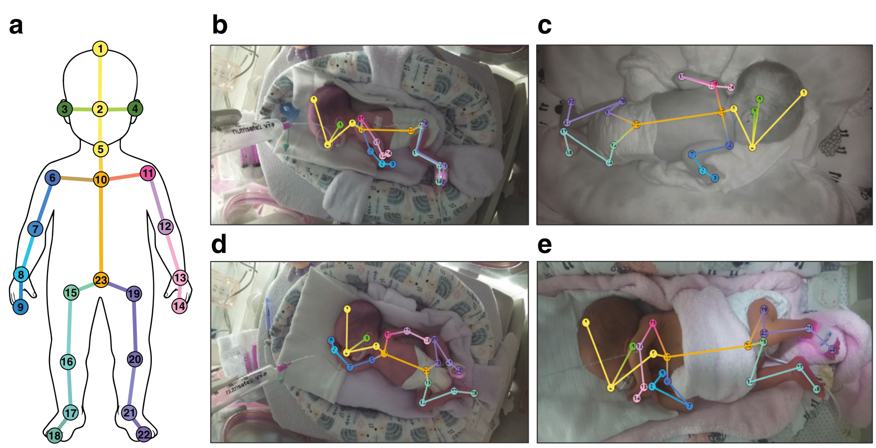

Fig. 1 Body pose model. a Illustration of the body keypoint locations used in the study. **b**–**e** Example views with body pose overlaid. Body alignments are right lateral (**b**), prone (**c**), supine (**d**), and semi-lateral (**e**), respectively.

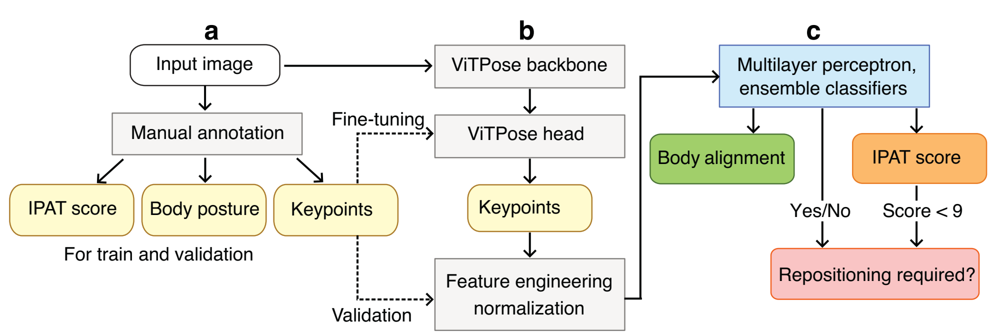

Fig. 2 Classification process flow chart. a Input images are manually annotated for IPAT score, posture, and body keypoints. **b** Keypoint annotations are used for augmenting and fine-tuning the ViTPose pose estimator. **c** Multiple classifier methods are trained and evaluated on manually annotated and estimated keypoints to provide classification of body alignment and repositioning indication. Suboptimal positioning requiring an intervention decision is derived in two different methods: using the thresholded regression estimated cumulative IPAT score and using binary prediction.

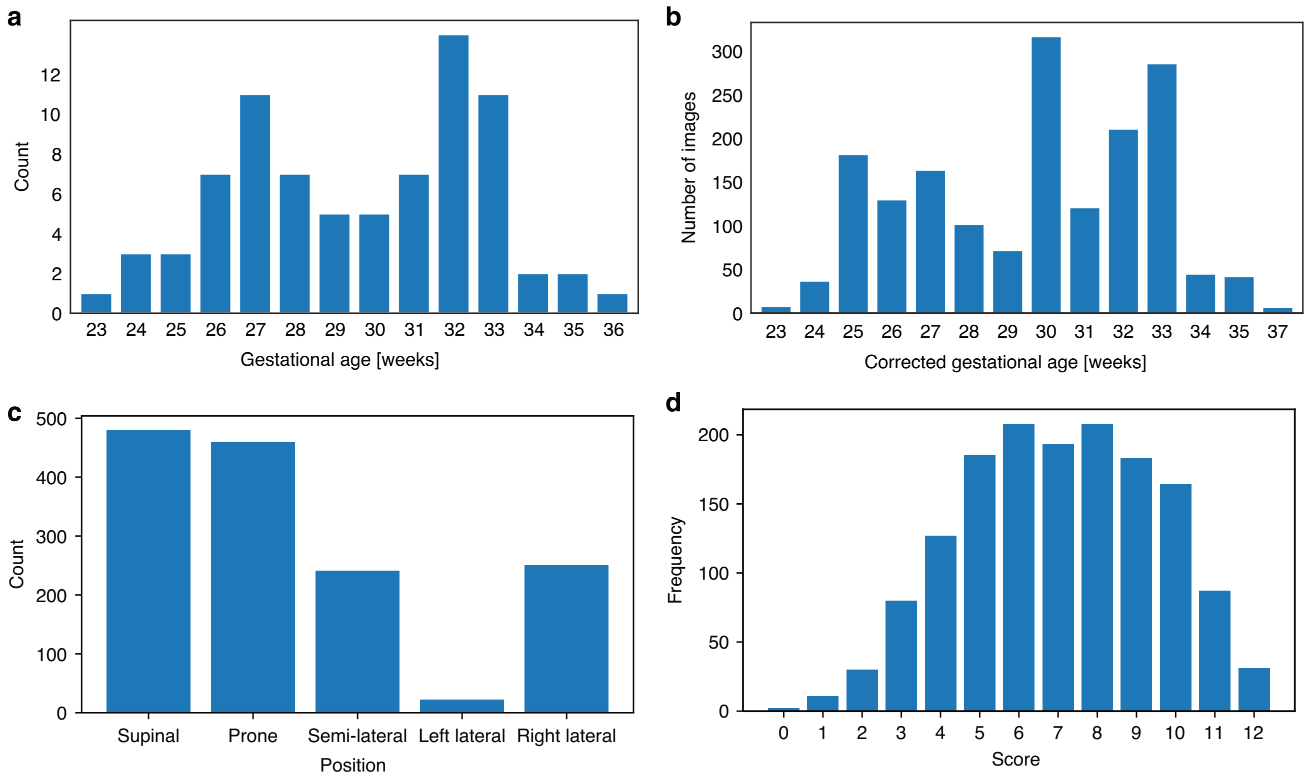

Fig. 3 Study population statistics. **a** Gestational age at birth for infants included in the dataset. **b** Number of images by infants' corrected gestational age at the time of recordings. **c** Frequency of classified body postures in the evaluated dataset. **d** IPAT accumulated scores for the evaluated dataset.

In the following sections, the manually annotated model was evaluated across classifier types and compared against the fine-tuned ViTPose model, isolating the contribution of the pose estimator from that of the downstream classifier.

RESULTS

Study population and data

Figure 3 depicts the population characteristics and the data distributions. Of 88 initially enrolled infants, 8 were excluded due to insufficient data. From the remaining 80 infants with recorded data, 50 met gestational age inclusion criteria and had sufficient high-quality video data for analysis. From 3385 initially gathered images, 1080 met all inclusion criteria, including adequate visibility of the body keypoints (≥70%), documented sleep state, and image quality thresholds. After applying adaptive sampling, the final dataset comprised 1510 images from the $n = 50$ infants, from both random surveillance (2-h/10 min intervals, $n = 46$, 1084 images) and longitudinal tracking protocols (10-min intervals, $n = 4$ infants, 426 images), providing both broad population coverage and temporal density for assessing positioning stability. Sleep stages were determined retrospectively from the video recordings by trained clinical staff using behavioral criteria. The cohort comprised 51.5% male and 48.5% female infants (sex assigned at birth). The mean gestational age at birth of 29.7 weeks (SD: 3.35 weeks, range: 23–36 weeks), and the analyzed images were captured at a mean postnatal age of 10.4 days (SD: 8.3 days). Their mean body length was 38.9 cm (SD: 5.5 cm), and their mean weight was 1297 g (SD: 536.9 g) at the start of recording. Of the included infants, 71.9% received non-invasive respiratory support during the monitoring period. Images per infant ranged from 5 to 192 (mean: 35.1, SD: 37.6), reflecting differences in sampling pool, length of stay, and image quality after applying exclusion criteria. Most images (75.8%) were captured during nighttime or in covered incubators using infrared illumination, resulting in more grayscale than color images. IPAT scores ranged from 0 to 12, with a near-symmetric, flat distribution around the mean (mean: 7.01, variance: 6.07). The longitudinal dataset has a similar IPAT score distribution (range: 1 to 12, mean: 7.41, variance: 5.93) to the complete dataset. Inter-annotator keypoint location variability averaged 0.0278 relative units of normalized body size (2.2 cm in the original image plane), and 6.2% of the images were excluded due to incomplete keypoint count. Missing keypoint coordinates were encoded for each classifier as non-numeric values in the tree-based models and mean imputation for the SVM.

Performance of body alignment

In body alignment classification of the five alignment categories with both manually annotated and estimated keypoints, the ensemble classifiers performed better than the SVM model, especially the Random Forest classifier (Table 1). Three major categories were also created by merging lateral directions together and the prone with semi-lateral. Figure 4 illustrates the best-performing Random Forest classifier confusion matrix and feature importance map.

Performance of positioning assessment

An important goal of the positioning assessment is to catch deteriorating situations. In the IPAT framework, repositioning is recommended when the cumulative score is in the range of 0–8 from the 0–12 scale. Figure 5 shows the suboptimal positioning prediction as a function of cumulative IPAT scores. The prediction

Table 1. Results for body alignment classification by multiple classifiers with five classes (supine, prone, 45° semi-lateral, left lateral, and right lateral positions) and with three merged categories (supine, prone, and lateral) using a Random Forest classifier on leave-one-infant-out cross-validation data

	SVM	Random Forest	Random Forest Merged categories	XGBoost	CatBoost	End-to-end, fine-tuned ViTPose + Random Forest
Accuracy	0.39 (0.31, 0.48)[a]	**0.80 (0.74, 0.86)**	**0.90 (0.85, 0.93)**	0.76 (0.69, 0.81)	0.77 (0.71, 0.82)	**0.78 (0.72, 0.84)**
Precision	0.27 (0.19, 0.38)	**0.81 (0.75, 0.86)**	**0.88 (0.84, 0.93)**	0.75 (0.69, 0.80)	0.64 (0.61, 0.70)	**0.79 (0.73, 0.84)**
Recall	0.39 (0.31, 0.48)	**0.80 (0.74, 0.86)**	**0.89 (0.84, 0.93)**	0.74 (0.68, 0.80)	0.70 (0.68, 0.75)	**0.78 (0.73, 0.84)**
F1-score	0.30 (0.21, 0.39)	**0.79 (0.73, 0.85)**	**0.88 (0.83, 0.93)**	0.73 (0.67, 0.79)	0.61 (0.58, 0.64)	**0.77 (0.71, 0.83)**

The bold valued classifiers (columns) performed the best.
[a]Macro average (95% confidence intervals).

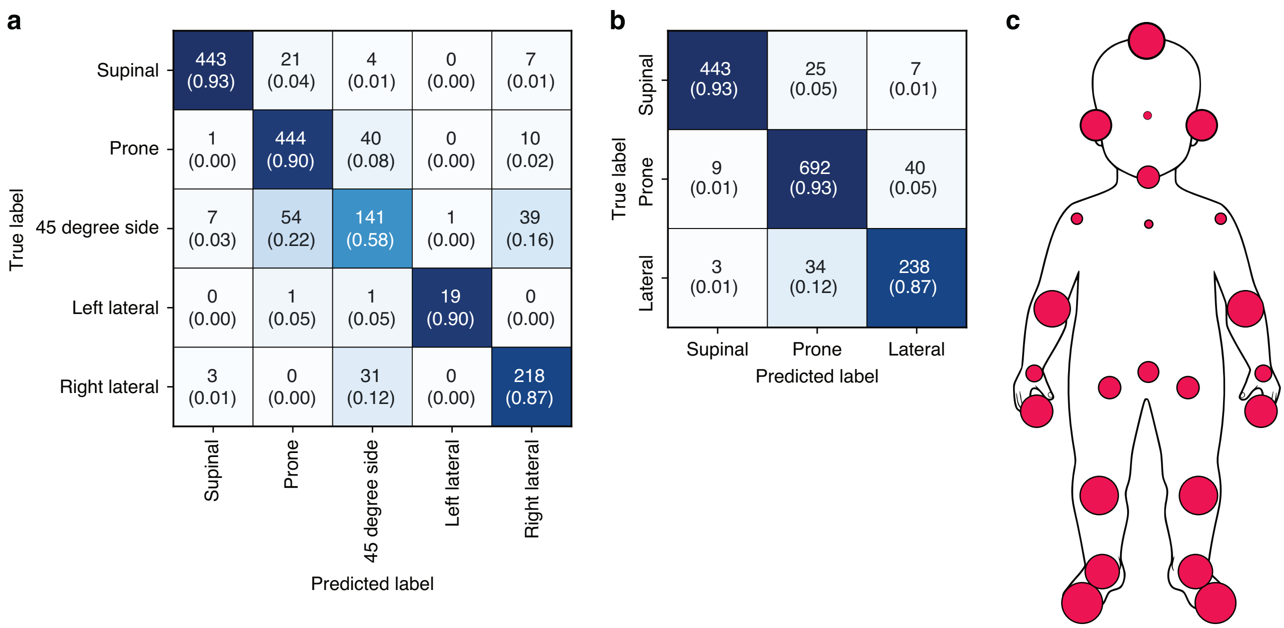

Fig. 4 Confusion matrix of body posture classification. Confusion matrices using the best-performing Random Forest classifier, **a** for five classes, as the result of leave-one-infant-out cross-validation, and **b** for three separately trained and evaluated significant alignment cases: a) supine, prone, and lateral. The numbers in the matrix are the raw counts with the normalized values in brackets. **c** Shapley beeswarm absolute value importance data mapped to keypoint locations, scaled for visibility.

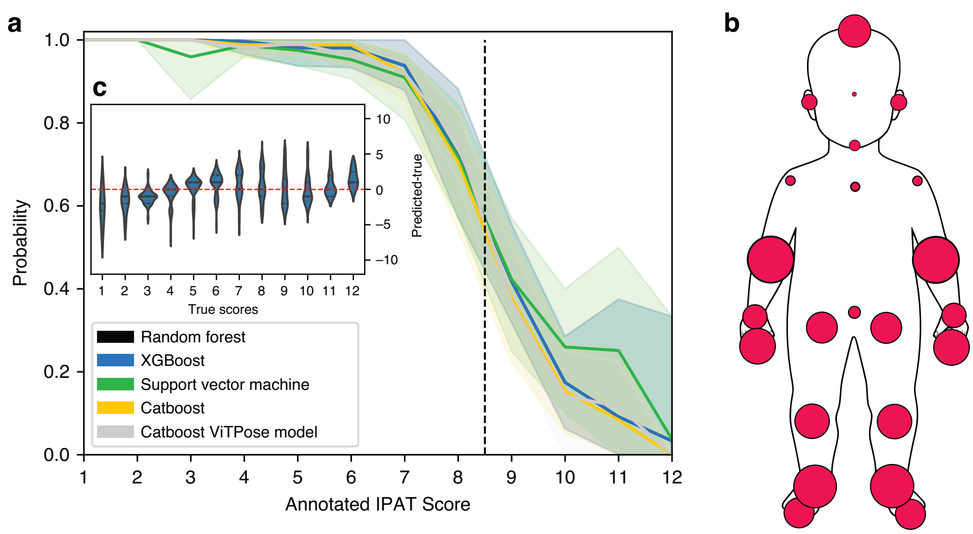

Fig. 5 Probability of the correct prediction of suboptimal positioning (IPAT Score < 9) as a function of the annotated IPAT Score with 95% confidence intervals. a XGBoost, SVM, Random Forest, and CatBoost evaluated on manually annotated keypoints, and a CatBoost classifier using the ViTPose pose estimator model. **b** Shapley beeswarm importance data mapped to keypoint locations, scaled for visibility. **c** Violin plot showing the distribution of the difference between the predicted and annotated individual scores.

Table 2. Classification and regression results for IPAT score classification on leave-one-infant-out cross- validation data

	SVM	Random Forest	XGBoost	CatBoost	End-to-end fine-tuned ViTPose + CatBoost
Precision	0.45	0.76	0.77	**0.76 (0.66, 0.78)**[a]	**0.77 (0.72, 0.82)**
F1-score	0.44	0.56	0.60	**0.63 (0.52, 0.78)**	**0.60 (0.54, 0.65)**
RMSE[b]	3.11	2.41	2.28	**2.42 (2.32, 3.30)**	**2.37 (2.08, 2.65)**
MAE[b]	1.31	1.15	1.12	**1.08 (0.98, 1.14)**	**1.12 (1.06, 1.17)**

The bold valued classifiers (columns) performed the best.
[a]Macro average (95% confidence intervals).
[b]Root mean square error (RMSE) and mean absolute error (MAE) of the test and the predicted cumulative scores.

Table 3. Macro average results for body repositioning indication on leave-one-infant-out cross-validation data for two strategies

	SVM	Random Forest	XGBoost	CatBoost	End-to-end, fine-tuned ViTPose + CatBoost
Accuracy	0.86/0.85[a]	0.86/0.87	0.88/0.87	0.89/**0.87**	0.88/**0.87**
Precision	0.91/0.71	0.91/0.76	0.92/0.75	0.93/**0.77**	0.93/**0.71**
Specificity	0.92/0.94	0.65/0.95	0.71/0.95	0.72/**0.95**	0.92/**0.94**
Sensitivity	0.92/0.51	0.92/0.59	0.92/0.57	0.92/**0.57**	0.73/**0.59**
F1-score	0.91/0.59	0.91/0.66	0.92/0.65	0.94/**0.65**	0.92/**0.64**
Fleiss's κ	0.57/0.50	0.57/0.59	0.64/0.57	0.65/**0.57**	0.63/**0.56**

The bold valued classifiers (columns) performed the best.
Strategy a: detailed body positioning scores are predicted by regression, and the cumulative score is thresholded; Strategy b: repositioning indication is predicted directly as a binary output with no detailed assessment.
[a]Strategy a/Strategy b.

shown is the probability that the thresholded sum of the individually predicted IPAT scores correctly estimates the required logic value. Figure 5 also depicts the body-mapped importance map of the best-performing CatBoost classifier. Tables 2–4 list the evaluation metrics of the detailed score prediction.

To directly capture body misalignment, we also trained the evaluated models for this binary decision with no refined scoring. Tables 3 and 4 present detailed statistics for the model comparison, along with regression-based results for the cumulative IPAT score. Figure 6 shows the AUROC and AUPRC for the binary decision classifiers.

DISCUSSION

The results demonstrate that automated video-based assessment of infant positioning is feasible. The Random Forest classifier reached better performance for posture classification, while ensemble classifiers (Random Forest, XGBoost, CatBoost) achieved a similar performance for predicting repositioning indication. Though 3D capture techniques demonstrate superior body pose estimation,[19,24,25] we evaluated a more broadly deployable 2D video solution compatible with existing NICU infrastructure. Notably, 75.8% of the analyzed images were captured under nighttime or infrared conditions; most published neonatal pose-estimation work relies predominantly on daytime RGB imagery, whereas continuous clinical monitoring requires robust performance across lighting regimes.

Model performance

The system achieves high accuracy in classifying both five-category body alignment and major alignment categories using the Random Forest classifier. This accurate alignment tracking enables timely decision support for repositioning infants at 2–3-h intervals according to best practice guidelines, which helps prevent pressure-related ulcers and other complications. The use of the end-to-end automatic solution did not significantly affect classification performance, demonstrating robust decision support capabilities for tracking infant posture.

Individual IPAT component regression (Strategy “a”), the models showed moderate performance with wider confidence intervals. The most challenging categories were shoulder retraction/rounding and hip/pelvis positioning (F1-Score ~ 0.50), likely reflecting assessment uncertainty. Best-performing categories were head rotation and hand positioning, representing more visually distinct features. Compared to direct binary prediction for repositioning classification (Strategy “b”), the component-regression-based approaches achieved higher AUPRC but lower AUROC, resulting in higher sensitivity with lower specificity for all models (Table 4).

Generally, all ensemble classifiers achieved strong and similar performance accuracy, falling into the 85–87% range with tight confidence intervals, indicating this is a classification task well-suited to these algorithms. However, the moderate inter-annotator agreement suggests that annotation noise represents the primary limiting factor for the performance. The models replicate expert consensus annotation rather than an outcome-anchored gold standard.

Model robustness and interpretability

SHAP analysis revealed which anatomical landmarks most influenced decisions, enhancing clinical interpretability and body model selection. By using the modified ViTPose’s keypoint estimator, the performance dropped only slightly. This indicates that fine-tuning of the modified output structure is reliable and requires additional samples. The robustness of the ViTPose backbone, trained on large-scale general image datasets, showed advantages compared to the more specialized infant-centric AGMA-PESS model. Notably, the automated pipeline occasionally outperformed manual keypoint annotation. We attribute this to instances where annotators omitted certain keypoints, whereas the automatic estimator provided reasonable predictions for these missing landmarks.

Clinical implementation and cultural challenges

Real-world deployment faces technical complexities: detecting patient presence, handling lighting changes, identifying active

Table 4. Area Under Receiver Operating Characteristic curve and Area Under Precision-Recall Curve mean values and confidence intervals for the binary output body repositioning indication classification on leave-one-infant-out cross-validation data

	SVM	Random Forest	XGBoost	CatBoost	End-to-end, fine-tuned ViTPose + CatBoost
AUROC	0.81 (0.78, 0.83)/0.89 (0.81, 0.95)[a,b]	0.84 (0.83, 0.85)/0.86 (0.79, 0.92)	0.83 (0.81, 0.85)/0.86 (0.82,0.90)	0.83 (0.81, 0.84)/**0.86 (0.79, 0.87)**	0.82 (0.79, 0.88)/**0.87 (0.84, 0.92)**
AUPRC	0.91 (0.90, 0.92)/0.67 (0.58, 0.73)	0.92 (0.91, 0.93)/0.69 (0.59, 0.78)	0.92 (0.91, 0.92)/0.67 (0.59, 0.73)	0.92 (0.91, 0.93)/**0.69 (0.59, 0.79)**	0.92 (0.90, 0.93)/**0.70 (0.61, 0.82)**

Strategy a: detailed body positioning scores are predicted by regression, and the cumulative score is thresholded; Strategy b: repositioning indication is predicted directly as a binary output with no detailed assessment.
The bold valued classifiers (columns) performed the best.
[a]Strategy a/Strategy b.
[b]Macro average (95% confidence intervals).

intervention periods, managing occlusion from equipment and positioning aids, and privacy considerations.

The Strategy “a” is more appropriate for clinical deployment due to superior sensitivity and F1-score, minimizing missed clinically relevant events, while the Strategy “b” is better suited for confirmatory workflow applications requiring higher specificity. Additionally, the low computational requirements, needing only a few minutes periodic rather than real-time assessment, enable deployment in resource-limited settings.

Substantial variation exists in neonatal care practices across institutions, countries, and cultures,[26–28] limiting visual assessment, a problem that automation does not resolve. In parts of Asia, Central Europe, and the Mediterranean, swaddling and nesting practices are more prevalent, often resulting in partial or complete body coverage.[26] Conversely, in the United Kingdom, the United States, Australia, and Scandinavia, guidelines typically promote uncovered positioning, emphasizing spontaneous movement.[26] These regional variations create challenges for visual monitoring, positioning aids, and swaddling obscure anatomical landmarks needed for accurate assessment.

Any clinical deployment of continuous video monitoring in the NICU must address infant and family data protection from the outset. A key design choice was that the inference pipeline operates on derived anatomical keypoints rather than raw video, allowing long-term monitoring, audit trails, and analytics to be performed on non-identifying skeletal data while raw recordings are retained only transiently. Because keypoint extraction and classification are computationally modest, on-premise closed, and secure deployment is feasible for standard infrastructure and is not expected to become a limiting factor.

Limitations

Key limitations include retrospective design with post-hoc image selection, two-center validation within a single institution system, ≥70% body visibility requirement excluding heavily swaddled infants, lack of extensive testing during active caregiving, and absence of long-term outcome data linking positioning scores to developmental endpoints. Notably, inter-annotator agreement placed an upper bound on model performance, indicating that classifier capability was not the limiting factor. This motivates the adoption of modified positioning assessment scoring systems better suited to real-world clinical variation,[14,15,29] which could improve both annotation consistency and a path toward validating objective clinical endpoints.

Future directions

The applied body visibility constraint means heavily swaddled infants may be inadequately monitored, limiting generalizability in certain cultural contexts, or more experimental, 3D, and distance mapping imaging technology is required. Generalization of the model to more acute phases and open beds requires separate datasets and validation.

CONCLUSIONS

We have demonstrated that automated assessment of the body position of infants cared for in NICU is feasible using video-based machine learning. Our proof-of-concept system can serve as a basis for the development of continuous objective body position monitoring in the NICU. We demonstrated that the ensemble classifiers achieved statistically indistinguishable performance; the choice among them can be made on practical grounds such as training time, inference cost, and ease of deployment. However, cultural differences in swaddling and nursing practices and requirements for body visibility pose significant implementation challenges. Therefore, further multi-center studies are needed for further development and future clinical application. A prospective interventional study linking continuous

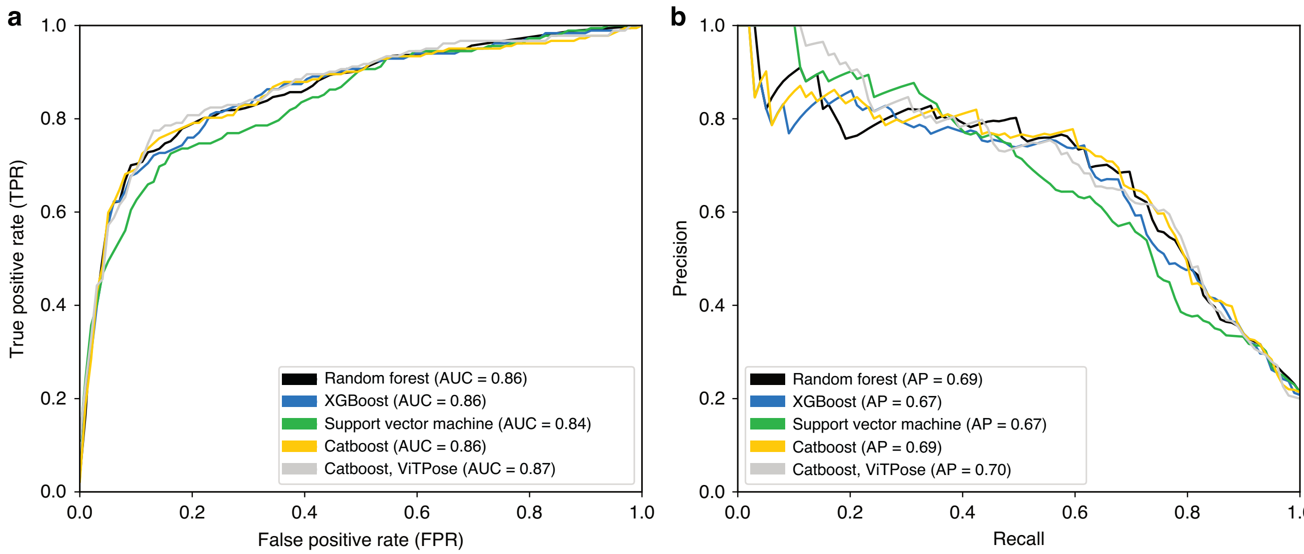

Fig. 6 Prediction metrics of suboptimal positioning. Receiver Operating Characteristic curves (**a**) and Precision-Recall curves (**b**) for predicting required repositioning by the binary output classifiers using leave-one-infant-out cross-validation. AUROC and AUPRC values are reported as macro averages in the figure legends.

automated IPAT monitoring to clinical endpoints is also a prerequisite for clinical adoption.

DATA AVAILABILITY

The datasets and program codes generated and analyzed during the current study are not publicly available due to institutional data-sharing policy, but are available from the corresponding author upon reasonable request.

REFERENCES

1. Chanpura, D., Mukkamala, N. & Gupta, N. Different positioning practices in neonatal intensive care unit and their effects on neuromotor development of preterm neonates: a systematic review. *J. Clin. Diagn. Res.* **19**, 36 (2025).
2. Prescott, M. G. et al. Infant positioning for promoting development and preventing morbidity in preterm infants. *Cochrane Database Syst. Rev.* **2024**, CD016029 (2024).
3. Butler, A. S. & Behrman, R. E. (eds) *Preterm Birth: Causes, Consequences, and Prevention* (National Academy Press, 2007).
4. Shepherd, K. L. et al. When does prone sleeping improve cardiorespiratory status in preterm infants in the NICU? *Sleep* **43**, zsz256 (2020).
5. Ballout, R. A., Foster, J. P., Kahale, L. A. & Badr, L. Body positioning for spontaneously breathing preterm infants with apnoea. *Cochrane Database Syst. Rev.* **1**, CD004951 (2017).
6. McKinney, C. M., Holt, V. L., Cunningham, M. L., Leroux, B. G. & Starr, J. R. Maternal and infant characteristics associated with prone and lateral infant sleep positioning in Washington state, 1996–2002. *J. Pediatr.* **153**, 194–198 (2008).
7. Charafeddine, L. et al. Targeted educational program improves infant positioning practice in the NICU. *Int. J. Qual. Health Care* **30**, 642–648 (2018).
8. Westrup, B. Newborn Individualized Developmental Care and Assessment Program (NIDCAP)—family-centered developmentally supportive care. *Early Hum. Dev.* **83**, 443–449 (2007).
9. Warren, I., Mat-Ali, E., Green, M. & Nyathi, D. Evaluation of the Family and Infant Neurodevelopmental Education (FINE) programme in the UK. *J. Neonatal Nurs.* **25**, 93–98 (2019).
10. Coughlin, M., Lohman, M. B. & Gibbins, S. Reliability and effectiveness of an infant positioning assessment tool to standardize developmentally supportive positioning practices in the neonatal intensive care unit. *Newborn Infant Nurs. Rev.* **10**, 104–106 (2010).
11. Ar, I. & Balcı, S. The effect of the use of positioning materials and hand-face maneuvers on physiologic parameters and stress symptoms in preterm newborns. *Ital. J. Pediatr.* **51**, 239 (2025).
12. Yun, E. J. & Kim, T. I. Development and effectiveness of an educational program on developmental positioning for neonatal intensive care unit nurses in South Korea: a quasi-experimental study. *Child Health Nurs. Res.* **28**, 70 (2022).
13. Mugisha, S., Kisitu, R., Komakech, F. & Favor, E. Vision-based embedded system for noncontact monitoring of preterm infant behavior in low-resource care settings. Preprint at https://arxiv.org/abs/2509.02018 (2025).
14. Grafton, A. et al. Neonatal pose estimation in the unaltered clinical environment with fusion of RGB, depth and IR images. *npj Digit. Med.* **8**, 539 (2025).
15. Hesse, N. et al. Body pose estimation in depth images for infant motion analysis. In *Proc. 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 1909–1912 (IEEE, 2017).
16. Soualmi, A. et al. AGMA-PESS: a deep learning-based infant pose estimator and sequence selector software for general movement assessment. *Front. Pediatr.* **12**, 1465632 (2024).
17. Groos, D., Adde, L., Støen, R., Ramampiaro, H. & Ihlen, E. A. Towards human-level performance on automatic pose estimation of infant spontaneous movements. *Comput. Med. Imaging Graph.* **95**, 102012 (2022).
18. Cao, X. et al. AggPose: deep aggregation vision transformer for infant pose estimation. In *Proc. 2022 Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)* 5045–5051 (2022).
19. Liu, J., Wang, Y., Liu, Y., Xiang, S. & Pan, C. 3D PostureNet: a unified framework for skeleton-based posture recognition. *Pattern Recognit. Lett.* **140**, 143–149 (2020).
20. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. In Proc. *Advances in Neural Information Processing Systems* **30**, 4765–4774 (Curran Associates, Inc., 2017).
21. Olczak, J. et al. Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. *Acta Orthop.* **92**, 513–525 (2021).
22. Jahn, L. et al. Comparison of marker-less 2D image-based methods for infant pose estimation. *Sci. Rep.* **15**, 12148 (2025).
23. Xia, H. & Zhang, Q. VitPose: multi-view 3D human pose estimation with vision transformer. In *Proc. 2022 IEEE 8th International Conference on Computer and Communications (ICCC)* 1922–1927 (IEEE, 2022).
24. Moccia, S., Migliorelli, L., Carnielli, V. & Frontoni, E. Preterm infants' pose estimation with spatio-temporal features. *IEEE Trans. Biomed. Eng.* **67**, 2370–2380 (2019).
25. Addison, P. S. et al. Touchless monitoring of neonatal activity: a multi-center study. *Pediatr. Res.* https://doi.org/10.1038/s41390-025-03905-5 (2025).
26. Keles, E. & Bagci, U. The past, current, and future of neonatal intensive care units with artificial intelligence: a systematic review. *npj Digit. Med.* **6**, 220 (2023).
27. Deger, V. B. & Butun, A. Traditional care practices known and/or used by different ethnic groups for newborns during the postpartum period. *BMC Public Health* **25**, 1622 (2025).
28. Al-Sagarat, A. Y. & Al-Kharabsheh, A. Traditional practices adopted by Jordanian mothers when caring for their infants in rural areas. *Afr. J. Tradit. Complement. Altern. Med.* **14**, 1–9 (2017).
29. Cruz, I. R., Curado, M. A. S. & Brantes, A. L. G. Infant Positioning Assessment Tool: transcultural adaptation and validation for Portuguese preterm neonates. *J. Neonatal Nurs.* **29**, 669–674 (2023).

AUTHOR CONTRIBUTIONS

P.F. conceptualized and developed the study, analyzed, and interpreted the data. M.Sz. helped in designing the experimental protocol, in financial support, and in final approval. Z.L.R. cleaned the data, analyzed the statistical coverage, and helped draft and edit the manuscript. P.F. drafted and edited the manuscript. Á.N., I.J., and Á.A. assisted with model building, model training and validation, and interpreting the data. J.V. and F.F. managed organizing, maintaining, and documenting data, editing the manuscript, and aligning the study to clinical routine. All authors approved the final version of this manuscript.

FUNDING

This work was supported by Project No. RRF-2.3.1-21-2022-00011, titled National Laboratory of Translational Neuroscience, has been implemented with the support provided by the Recovery and Resilience Facility of the European Union within the framework of Programme Széchenyi Plan Plus. A.Á. is supported by the EKÖP-25-3 University Research Scholarship Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund. Open access funding provided by HUN-REN Institute for Computer Science and Control.

COMPETING INTERESTS

The authors declare no competing interests.

CONSENT STATEMENT

This study was conducted under the IRB approval SE-RKEB: 265/2022 of Semmelweis University, Hungary, with the written consent of parents.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Péter Földesy.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.