Global Health & Medicine. 2026; 8(3):227-232.

CORRESPONDENCE

DOI: 10.35772/ghm.2026.01061

Artificial intelligence (AI)-based pose estimation detects movements linked to unplanned tube removal in ICU patients

Aya Umeda[1,2,*], So Mizuno[3,4], Fumio Ishizaki[5], Tatsuya Okamoto[6]

[1] Department of Adult Health Nursing, National College of Nursing, Japan (NCNJ), Japan Institute for Health Security, Tokyo, Japan;
[2] Department of Nursing, Japan Institute for Health Security, Tokyo, Japan;
[3] Japan Institute of Life Design Counseling, Chiba, Japan;
[4] Division of Educational Sciences, Hiroshima University, Hiroshima, Japan;
[5] Modal Stage Inc, Tokyo, Japan;
[6] Department of Emergency and Critical Care Medicine, Japan Institute for Health Security, Tokyo, Japan.

Abstract: Unplanned removal of life-sustaining tubes in intensive care units (ICUs) poses serious risks, yet existing monitoring methods relying on physical restraints have ethical and clinical drawbacks. Here we applied artificial intelligence (AI)-based pose estimation using MediaPipe to analyze ICU surveillance videos, extracting skeletal coordinates to detect movements associated with tube removal. Using Singular Spectrum Transformation for change-point detection, we identified movement changes corresponding to tube-removal behaviors in three consented cases, achieving average precision values substantially above chance. These preliminary results demonstrate that AI-driven, contactless motion analysis can capture clinically relevant signals from existing ICU infrastructure without additional patient burden. Although limited by sample size and environmental factors, this approach holds promise for real-time, non-invasive monitoring to reduce reliance on physical restraints and enhance patient safety in critical care settings.

***Keywords*:** AI-based monitoring, ICU patient safety, pose estimation technology, unplanned tube removal detection, contactless patient monitoring

1. Introduction

Patients admitted to the intensive care unit (ICU) often depend on life-sustaining devices such as endotracheal tubes, nasogastric tubes, arterial lines, and various drainage catheters. Unplanned removal of these devices is a serious clinical problem, as it can immediately threaten patient survival. For example, approximately 40–60% of patients who remove their endotracheal tube require reintubation, which is associated with prolonged mechanical ventilation, longer ICU and hospital stays, and increased healthcare costs (*1*). Although the incidence of unplanned endotracheal tube removal is relatively low—ranging from 0.05% to 2% among mechanically ventilated ICU patients (*1*)—its consequences can be severe. A pooled prevalence of about 6.7% has been reported across all ICU patients receiving mechanical ventilation (*2*), making reliable prevention a critical clinical priority.

Physical restraints have been widely adopted in ICUs worldwide, with restraint rates reported to range from 8.7% to 59.1% (*3*). However, restraint carries significant ethical and clinical concerns: beyond violating patient autonomy, it is associated with focal skin injury, delirium, post-traumatic stress disorder (PTSD), and other sequelae of post-intensive care syndrome (PICS) (*2,4,5*). Furthermore, restraint does not reliably prevent device removal; even among restrained patients, 61.4% still succeeded in removing their devices (*6*). These findings underscore the urgent need for a fundamentally different approach—specifically, proactive, real-time detection of tube-removal behaviors before they occur. This need highlights the unresolved dilemma faced by clinicians: preventing life-threatening events while also avoiding harms inherent in prolonged restraint.

To our knowledge, no published study has prospectively characterized movement patterns that immediately precede unplanned tube removal in ICU patients, leaving open the question of whether such behaviors are detectable in real time. Since 2020, our group has systematically investigated approaches to quantify these movements with the goal of developing a contactless early warning system. We found that optical motion capture was incompatible with the ICU environment. Inertial motion capture (IMC), while promising for three-dimensional movement data, proved unreliable in supine patients due to posterior sensor occlusion by the mattress, and also posed challenges related to sensor attachment in critically ill patients. These limitations led us to explore artificial intelligence (AI)-based pose estimation—a contactless approach that reconstructs body landmark positions from standard surveillance footage without need for wearable devices. This builds on our preliminary findings first presented at a national conference (*7*).

2. Real-world ICU video data and ethical challenges

This prospective observational study was conducted between 2021 and 2023 at the ICU of a tertiary hospital in Tokyo, Japan. Security camera footage was reviewed for unplanned tube removal events as they occurred during the study period, and patients who experienced tube removal were approached for consent.

Obtaining informed consent proved exceptionally challenging throughout the study period. Many patients had pre-existing cognitive impairment or dementia, or were in a critical condition from which they did not survive, rendering direct consent impossible. Furthermore, the study overlapped substantially with the COVID-19 pandemic, during which family visitation was severely restricted at the institution, making surrogate consent from family members equally difficult to obtain. As a result, valid consent was obtained from only 13 of the 58 patients who experienced unplanned tube removal during the three-year study period, underscoring the exceptional difficulty of enrolling this patient population in real-world ICU research. Characteristics and imaging conditions of all 13 video cases are summarized in Table 1.

From each consented case, video frames were extracted centered on the time of the tube removal event, using a maximum window of 5 minutes per clip. Videos were recorded at 5 frames per second (fps) with a resolution of 640 × 480 pixels. Facial blurring was applied to video frames used in publication figures to protect patient privacy. The study was approved by ethics committee of National Center for Global Health and Medicine (approval No. 004011), and written informed consent was obtained from all participants or their surrogates in accordance with the Declaration of Helsinki.

3. Pose estimation and change-point analysis of hand-to-face movements

Pose estimation was conducted using MediaPipe (Google LLC), an open-source machine learning framework that applies deep neural networks to extract three-dimensional (3D) coordinates for 33 anatomical landmarks from standard monocular video footage in real time (*8*). This approach enables continuous skeletal tracking—including the face, trunk, limbs, and hands—using existing surveillance cameras, without the need

Table 1. Characteristics and analyzability of ICU videos involving unplanned tube removal

No.	Tube type	Site	Pose est.	Presumed main reason for failure	Camera orientation	Lighting	Body visibility	SST	AP
1	Endotracheal tube	Oral	✓	Low ambient lighting (lights-out)	Frontal	Adequate	Full body	✓	0.65
2	Arterial line	L. radial	✗	—	Frontal	Inadequate	Upper body	✗	—
3	Nasogastric tube	Nasal	✓	—	Frontal	Adequate	Upper body	✓	0.64
4	Nasogastric tube	Nasal	✓	Face obscured by right-rear angle and IV stand	Frontal	Adequate	Upper body	✓	0.32
5	Nasogastric tube	Nasal	✗	Poor facial visibility and limited body information	Right-rear oblique	Adequate	Upper body	✗	—
6	Endotracheal tube	Oral	✗	Low ambient lighting (lights-out)	Frontal	Adequate	Upper body	✗	—
7	Nasogastric tube	Nasal	✗	Low ambient lighting (lights-out)	Frontal	Inadequate	Upper body	✗	—
8	Nasogastric tube	Nasal	✗	Low pose discriminability (tube removal pose in 88% of frames)	Frontal	Inadequate	Full body	✗	—
9	Nasogastric tube	Nasal	✗	Low ambient lighting (lights-out)	Frontal	Adequate	Upper body	✗	—
10	Nasogastric tube	Nasal	✗	Low ambient lighting (lights-out)	Frontal	Inadequate	Upper body	✗	—
11	Arterial line	L. radial	✗	Facial landmarks obscured by surgical mask	Frontal	Inadequate	Full body	✗	—
12	Nasogastric tube	Nasal	✗	Lateral camera angle	Frontal	Adequate	Full body	✗	—
13	Nasogastric tube	Nasal	✗		Right lateral	Adequate	Full body	✗	—

Note: Videos 2, 7, 8, 10, and 11 occurred during nighttime hours after lights-out, accounting for all cases with inadequate lighting. Failure reasons represent the most plausible explanation based on visual inspection; the exact cause could not be determined algorithmically. *Abbreviations*: ICU, intensive care unit; L, left; Pose est, MediaPipe-based pose estimation; SST, Singular Spectrum Transformation; AP, Average Precision; —, not applicable.

for additional sensors or modifications to the clinical environment.

MediaPipe was applied frame-by-frame to each extracted video clip, and the resulting skeletal coordinate data were visually inspected to assess estimation quality. For each video, a trained observer reviewed individual frames and evaluated whether the estimated landmark positions corresponded accurately to the patient's actual body landmarks visible in the footage. Cases in which landmarks were completely undetected or their estimated positions were judged to deviate substantially from the patient's actual posture—due to occlusion, poor lighting, or non-frontal camera angle—were classified as invalid and excluded from further analysis. Valid pose estimation was achieved in 3 of the 13 video clips (23.1%) (Table 1). Of 33 detectable landmarks, nine were selected for analysis: six body landmarks (including both thumbs and mouth corners) and three facial landmarks (nose and mouth corners), as these were considered most relevant to postures associated with tube removal.

For the three analyzable cases—one endotracheal tube removal (Video 1, oral) and two nasogastric tube removals (Videos 3 and 4, nasal)—six Euclidean distances between anatomically relevant landmark pairs were computed as time-series features. These pairs (left thumb–right mouth corner, left thumb–left mouth corner, right thumb–right mouth corner, right thumb–left mouth corner, nose–right mouth corner, and nose–left mouth corner) were chosen for their sensitivity to hand-to-face movements associated with both oral and nasal tube removal, resulting in a six-dimensional time-series dataset for each video.

To examine whether coordinates obtained *via* MediaPipe could capture tube-removal-related movement changes, we performed time-series change-point detection using the Singular Spectrum Transformation (SST). SST is a statistical method that detects shifts in distribution of sequential data by comparing trajectory and test matrices using singular value decomposition (*9*). For each video, SST parameters (window size w and lag L) were manually optimized after reviewing the data to align the peak change score with visually confirmed onset of tube-removal behavior. All computations were performed in Python (Version 3.11) on the AI Bridging Cloud Infrastructure (ABCI) of the National Institute of Advanced Industrial Science and Technology.

Given the severe class imbalance between normal frames and the few frames corresponding to tube-removal-related movements (approximately 10 abnormal frames per video), we used Average Precision (AP) as evaluation metric. AP quantifies how well tube-removal-related frames are correctly identified among the predominantly normal frames. To assess whether the SST-derived scores reflected meaningful signal above chance, AP values were also computed using random score distributions of equal length as a baseline. Absence of negative-control videos (footage from patients who did not attempt tube removal) is acknowledged as a limitation of the current study design.

4. Preliminary detection of tube-removal-related movement signals

Of the 13 consented video clips, valid skeletal coordinate time-series data were obtained from 3 cases (23.1%), comprising one endotracheal tube removal (Video 1) and two nasogastric tube removals (Videos 3 and 4). Failure in the remaining 10 cases was attributable to low ambient lighting after lights-out (Videos 2, 7, 8, 10, 11), non-frontal camera orientation (Videos 5, 13), surgical mask use obscuring facial landmarks (Video 12), poor facial visibility and limited body information (Video 6), and low pose discriminability due to tube removal posture being present in 88% of frames (Video 9) (Table 1). Among the three analyzable cases, all had adequate ambient lighting; one showed the patient's full body (Video 1) and two showed upper body only (Videos 3 and 4), with all patients clearly visible and centrally positioned within the frame.

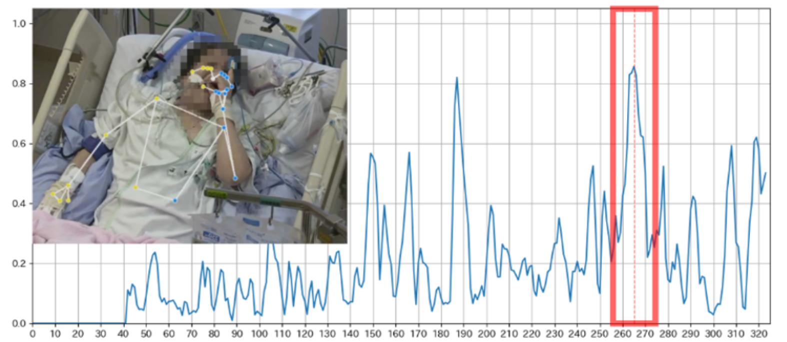
SST applied to the MediaPipe-derived inter-landmark distance data yielded prominent peaks in the change score at or immediately before the visually confirmed onset of tube-removal-related movements in all three analyzable cases (Figure 1). In Video 1 (endotracheal tube), peak change scores of a = 0.82 and a = 0.83 were observed at frames 188 and 265, corresponding to the left hand grasping the tube at the mouth and subsequent withdrawal movement, respectively. In Video 3 (nasogastric tube), consecutive peaks were observed from frame 750 through frame 1,230, corresponding to a sequence of actions including right-hand grasping and withdrawal of the tube, holding the withdrawn tube, reaching movements, and re-contact with the slack tube. In Video 4 (nasogastric tube), a peak of a = 0.65 at frame 1025 corresponded to left-hand tape removal followed by tube withdrawal.

As an exploratory evaluation, AP values for Videos 1 (AP = 0.65) and 3 (AP = 0.64) substantially exceeded the random-baseline AP (AP < 0.10 in all cases), indicating that detected change-score peaks reflected tube-removal-related signals rather than chance variation. In Video 4, the AP was 0.32, which exceeded the random baseline but was lower than the other two cases. This lower value was likely due to large bilateral arm movements occurring after tube removal, which generated higher change scores than the removal onset itself. These results are preliminary and exploratory, given the small sample size and absence of prospective validation.

5. Implications and limitations for future ICU monitoring

The present findings provide preliminary feasibility evidence that AI-based pose estimation using MediaPipe

(A): Video 1 (Endotracheal tube removal, oral)

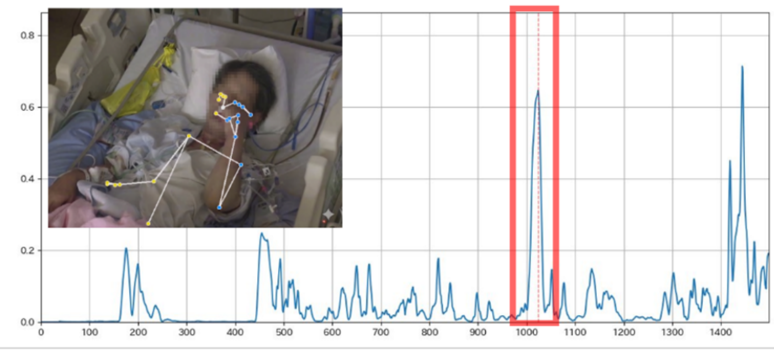


Frame	Degree of change	Motion description
188	0.82	Left-hand grasping of tube at mouth
265	0.83	Left-hand grasping and withdrawal of tube at mouth

(B): Video 3 — Nasogastric tube removal (nasal)

Frame	Degree of change	Motion description
750	0.26	Right-hand grasping and withdrawal of tube at nose
863	0.22	Right-hand holding of withdrawn tube
922	0.23	Right hand at chest, left hand reaching forward-left to grasp object
1230	0.22	Right-hand re-contact with slack of withdrawn tube

(C): Video 4 — Nasogastric tube removal (nasal)



Frame	Degree of change	Motion description
1025	0.65	Left-hand removal of tape around nose followed by tube withdrawal
1445	0.71	Bilateral hand contact with oxygen inhalation tube at nose

Figure 1. SST-based change scores with representative MediaPipe skeletal landmark overlays showing tube-removal-related movement onset in three analyzable ICU videos. (A) Video 1 (endotracheal tube removal, oral); **(B)** Video 3 (nasogastric tube removal, nasal). **(C)** Video 4 (nasogastric tube removal, nasal). In each panel, the blue line represents the SST-based change score at each frame (x-axis: frame number; y-axis: change score). The red box indicates primary SST peak region, which corresponded temporally to tube-removal-related movement observed in the video; the inset photograph shows the corresponding video frame as visual evidence of the detected movement. The dashed red line marks the frame of the primary change-score peak. The table within each panel lists frame number, change score, and corresponding motion description at key detected events. Facial blurring has been applied to all video frames shown for publication to protect patient privacy. *Abbreviation*: ICU, intensive care unit; SST, Singular Spectrum Transformation.

can extract skeletal coordinate data from ICU surveillance footage, and that SST-based change-point detection applied to these data can capture movement signals associated with tube-removal-related behaviors in real clinical ICU videos. To our knowledge, this is among the first demonstrations of such an approach using actual consented clinical footage of real tube removal events.

The potential of image-based motion analysis for unplanned tube removal detection is increasingly recognized internationally. An approach combining YOLOv3-based region-of-interest extraction with optical flow and support vector machine classification provided an early demonstration of video-based detection from RGB surveillance footage (*10*), and the AURA system subsequently used text-to-video diffusion models to generate synthetic ICU footage for training pose estimation-based detection of hand-to-tube proximity and agitation (*11*). The present study complements these approaches by using actual consented clinical footage, providing direct clinical validity that synthetic data cannot replicate. More broadly, Gabriel *et al.* demonstrated large-scale AI-driven patient monitoring across 11 hospitals (*12*), Nahin *et al.* showed deep learning pose estimation from infrared images in bedridden patients (*13*), and Feng *et al.* reported significant correlations between a pose-derived Movement Index (λMI) and GCS and RASS scores in ICU patients (*14*)—collectively affirming the clinical relevance of continuous posture-based monitoring.

Several limitations of the current work must be acknowledged in the interest of transparency. The proportion of videos yielding valid pose data was low (3/13, 23%), attributable to heterogeneous camera placement, low ambient lighting, non-frontal camera angles, and patient-specific factors such as surgical mask use. SST parameters were optimized post-hoc for each video, and absence of negative-control videos from patients who did not attempt tube removal precluded formal assessment of specificity. Notably, in one case (Video 9), the tube removal posture was present in 88% of frames, which likely prevented SST from detecting a meaningful change point. This highlights that change-point detection methods require sufficient contrast between normal and abnormal frames to function effectively. Additional limitations include small sample size, restriction to a single institution, lack of distinction between tube-removal-related and routine movements, and absence of prospective real-time validation. These limitations are particularly relevant given the challenging conditions of real-world ICU care during a pandemic.

Critically, the present results provide preliminary evidence that the core concept is feasible: AI-based pose estimation can capture anatomically meaningful movement signals from existing ICU surveillance infrastructure without modification to the clinical environment or additional burden on patients or staff. Practical barriers identified in this study are primarily technical and environmental, such as camera placement and lighting. Camera standardization protocols have already been demonstrated as feasible in large-scale clinical deployments (*12*), and use of infrared and RGB-D depth cameras has shown promise for reliable operation in low-light environments, including for pose estimation in bedridden patients (*13*). These technologies would directly address the nighttime lighting problem that accounted for most failures in this study, which is particularly relevant given that unplanned tube removal occurs more frequently during night shifts when lighting is reduced (*15,16*).

Looking ahead, a prospective implementation of this approach—with standardized camera placement, infrared imaging, and real-time alert functionality—could enable nursing staff to receive timely notifications before tube removal occurs, allowing targeted intervention without need for continuous physical restraint. As more annotated clinical data become available, supervised learning models could be developed to distinguish pre-removal movements from routine behaviors, advancing from statistical detection to genuine predictive monitoring. At the population level, continuous motion data could support individualized, dynamic risk stratification, replacing the current paradigm of default prolonged restraint with evidence-based, patient-specific care decisions.

The present study represents a first step toward that goal. Despite its limitations, it provides preliminary evidence from real clinical footage that AI-based contactless motion analysis may be a viable direction for ICU patient safety. Ultimately, a validated system of this kind could contribute to reducing unplanned tube removal and the physical and psychological burden of restraint in one of medicine's most vulnerable settings.

Funding: This work was supported by JSPS KAKENHI Grant Number JP.20K10753.

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Trenchat L, Galerneau LM, Ruckly S, *et al*. Self-extubation in critically ill patients: From the French OUTCOMEREA Network. Crit Care. 2025; 29:286.
2. Li P, Sun Z, Xu J. Unplanned extubation among critically ill adults: A systematic review and meta-analysis. Intensive Crit Care Nurs. 2022; 70:103219.
3. Zare-Kaseb A, Sarmadi S, Sanaie N, Emami Zeydi A. Prevalence and variability in use of physical restraints in intensive care units: A systematic review and meta-analysis. Aust Crit Care. 2025; 38:101210.
4. Franks ZM, Alcock JA, Lam T, Haines KJ, Arora N, Ramanan M. Physical restraints and post-traumatic stress disorder in survivors of critical illness. A systematic review and meta-analysis. Ann Am Thorac Soc. 2021; 18:689-697.

5. Acevedo-Nuevo M, Velasco-Sanz T, Del Olmo-Somolinos B, Vía-Clavero G. Ethical and legal considerations and recommendations for action in the physical restraint use in critically ill patients. Enferm Intensiva (Engl Ed). 2025; 36:100497.
6. Galazzi A, Adamini I, Consonni D, Roselli P, Rancati D, Ghilardi G, Greco G, Salinaro G, Laquintana D. Accidental removal of devices in intensive care unit: An eight-year observational study. Intensive Crit Care Nurs. 2019; 54:34-38.
7. Mizuno S, Ishizaki F, Umeda A, Okamoto T. Detection of self-extubation movements in the intensive care unit using a posture estimation model. Proc Annu Conf Jpn Soc Artif Intell. 2024; 38:3E5-GS-10-04.
8. Google AI for Developers. MediaPipe Pose Landmarker. *https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker* (accessed May 24, 2026).
9. Ide T, Sugiyama M. Anomaly Detection and Change Detection. Kodansha, Tokyo, Japan, 2015. (in Japanese)
10. Chen Y, Wang L, Wang G, Yang S, Wang Y, Xiang M, Zhang X, Chen H, Hu D, Cheng H. Spatio-temporal features for fast early warning of unplanned self-extubation in ICU. Eng Appl Artif Intell. 2024; 127:107294.
11. Seo J, Moon H, Jung KH, Oh N, Kim T. AURA: Development and validation of an augmented unplanned removal alert system using synthetic ICU videos. *https://arxiv.org/abs/2511.12241* (accessed May 24, 2026).
12. Gabriel P, Rehani P, Troy T, Wyatt T, Choma M, Singh N. Continuous patient monitoring with AI: Real-time analysis of video in hospital care settings. Front Imaging. 2025; 4:1547166.
13. Nahin SK, Acharjee S, Saha S, Das A, Hossain S, Haque MA. Human sleeping pose estimation from IR images for in-bed patient monitoring using image processing and deep learning techniques. Heliyon. 2024; 10:e36823.
14. Feng R, Richter F, Mari E, Gleason A, Le C, Kellner CP, Shrivastava RK, Fields M, Rapoport BI, Bederson JB, Schadt EE, Glicksberg BS, Richter F, Dangayach NS. Artificial intelligence monitoring of neurological status from patient videos in the neuroscience intensive care unit. Neurosurgery. 2026; doi: 10.1227/neu.0000000000003899.
15. Yeh SH, Lee LN, Ho TH, Chiang MC, Lin LW. Implications of nursing care in the occurrence and consequences of unplanned extubation in adult intensive care units. Int J Nurs Stud. 2004; 41:255-262.
16. Chang LC, Liu PF, Huang YL, Yang SS, Chang WY. Risk factors associated with unplanned endotracheal self-extubation of hospitalized intubated patients: A 3-year retrospective case-control study. Appl Nurs Res. 2011; 24:188-192.

Received May 27, 2026; Revised June 13, 2026; Accepted June 16, 2026.

Released online in J-STAGE as advance publication June 19, 2026.

**Address correspondence to*:
Aya Umeda, National College of Nursing, Japan (NCNJ), Japan Institute for Health Security, 1-2-1 Umezono, Kiyose-shi, Tokyo, 204-0024, Japan.
E-mail: a-umeda@umin.ac.jp