Continuous patient monitoring with AI: real-time analysis of video in hospital care settings

Paolo Gabriel*, Peter Rehani, Tyler Troy, Tiffany Wyatt, Michael Choma and Narinder Singh

Department of R&D, LookDeep Health, Oakland, CA, United States

Introduction: This study introduces an AI-driven platform for continuous and passive patient monitoring in hospital settings, developed by LookDeep Health. Leveraging advanced computer vision, the platform provides real-time insights into patient behavior and interactions through video analysis, securely storing inference results in the cloud for retrospective evaluation.

Methods: The AI system detects key components in hospital rooms, including individuals' presence and roles, furniture location, motion magnitude, and boundary crossings. Inference results are securely stored in the cloud for retrospective evaluation. The dataset, compiled with 11 hospital partners, includes over 300 high-risk fall patients and spans more than 1,000 days of inference. An anonymized subset is publicly available to foster innovation and reproducibility at lookdeep/ai-norms-2024.

Results: Performance evaluation demonstrates strong accuracy in object detection (macro F1-score = 0.92) and patient-role classification (F1-score = 0.98). The system reliably tracks the "patient alone" metric (mean logistic regression accuracy = 0.82 ± 0.15), enabling detection of patient isolation, wandering, and unsupervised movement-key indicators for fall risk and adverse events.

Discussion: This work establishes benchmarks for AI-driven patient monitoring, highlighting the platform's potential to enhance patient safety through continuous, data-driven insights into patient behavior and interactions.

KEYWORDS

artificial intelligence, medical imaging, computer vision, patient monitoring, RGB video, deep learning, healthcare analytics

1 Introduction

In hospitals, direct patient observation is limited–nurses spend only 37% of their shift engaged in patient care (Westbrook et al., 2011), and physicians average just 10 visits per hospital stay (Chae et al., 2021). This limited interaction hinders the ability to fully understand patient behaviors, such as how often patients are left alone, how much they move unsupervised, and how care allocation varies by time or condition. Virtual monitoring systems, which allow remote patient observation via audio-video devices, have improved safety, particularly for high-risk patients (Abbe and O'Keeffe, 2021).

Artificial Intelligence (AI) is transforming healthcare by enhancing diagnostic accuracy, streamlining data processing, and personalizing patient care (Davenport and Kalakota, 2019; Davoudi et al., 2019; Bajwa et al., 2021). While AI has found success in tasks like surgical assistance (Mascagni et al., 2022) and diagnostic imaging (Esteva et al., 2021), patient monitoring represents a critical frontier. Unlike these tasks, continuous patient monitoring involves real-time video analysis over extended periods, requiring AI systems

to process data efficiently and extract actionable insights spanning days, like day-over-day movement (Parker et al., 2022).

Continuous monitoring enhances safety and enables the detection of risks often missed during periodic assessments. For example, trends like delirium fluctuate throughout the day, but infrequent observations make these patterns hard to capture (Wilson et al., 2020). Similarly, patients occasionally leave their beds unattended—a key fall risk—yet monitoring every instance in real-time remains challenging. A robust computer vision-based system can provide immediate, context-aware insights into patient behavior (Chen et al., 2018), caregiver interactions (Avogaro et al., 2023), and room conditions (Haque et al., 2020). Such systems surpass traditional intermittent observation methods by detecting subtle patterns that inform care decisions (Lindroth et al., 2024).

However, achieving scalability, transparency, and adaptability in continuous monitoring systems presents significant challenges. These include efficiently processing video data at higher frame-rates (Posch et al., 2014), ensuring privacy compliance (Watzlaf et al., 2010), and adapting to dynamic hospital settings with varying lighting, camera angles, and patient behaviors. Addressing these technical and operational challenges is critical for AI-driven monitoring systems to gain acceptance and deliver meaningful outcomes, such as reducing falls and other preventable harms.

To bridge these gaps, this research presents a novel AI-driven system for continuous patient monitoring using RGB video (Figure 1), developed collaboratively with industry and healthcare providers. The LookDeep Health platform aims to enhance patient care by providing real-time monitoring and producing computer-vision-based insights into patient behavior, movement, and interactions with healthcare staff.

This study offers several key contributions:

1. **Implementation of advanced computer vision models:** our system utilizes state-of-the-art models for real-time predictions, including localization of people and furniture, monitoring boundary crossings, and calculating motion scores.
2. **Real-world validation:** we rigorously evaluated the system's performance in live hospital settings, illustrating its capability to present care providers with accurate data from continuous monitoring, and laying the foundation for future AI-enabled patient monitoring solutions.
3. **Dataset development:** we developed a comprehensive dataset encompassing over 300 high-risk fall patients tracked across 1,000 collective days and 11 hospitals, creating a valuable resource for studying patient behavior and hospital care patterns. This dataset is publicly available for further research at https://github.com/lookdeep/ai-norms-2024.

2 Methods

2.1 Study design

The LookDeep Health patient monitoring platform was deployed across 11 hospitals in three states within a single healthcare network. The system provides continuous, real-time monitoring of high-risk fall patients. Data collection adhered to institutional guidelines and patient consent procedures (see *Research Ethics*).

2.1.1 Participants

Patients monitored by LookDeep Health were primarily high-risk fall patients identified through mobility assessments as part of standard care protocols. This classification often results in the patient also being categorized as non-ambulatory during the inpatient stay (Capo-Lugo et al., 2023).

Data was organized into three subsets:

1. **Single-frame analysis:** periodic samples from monitoring sessions were used for training and testing object detectors, with over 40,000 frames collected to date. Only patients monitored during the first week of each month were included in the test set, providing 10,000 frames held out for consistent model evaluation.
2. **Observation logging:** ten patients who experienced falls were selected for additional annotation over a twelve month period (Figure 2A).
3. **Public dataset:** over 300 high-risk fall patients were monitored during a six month period, excluding those monitored for less than two days (Figure 2B).
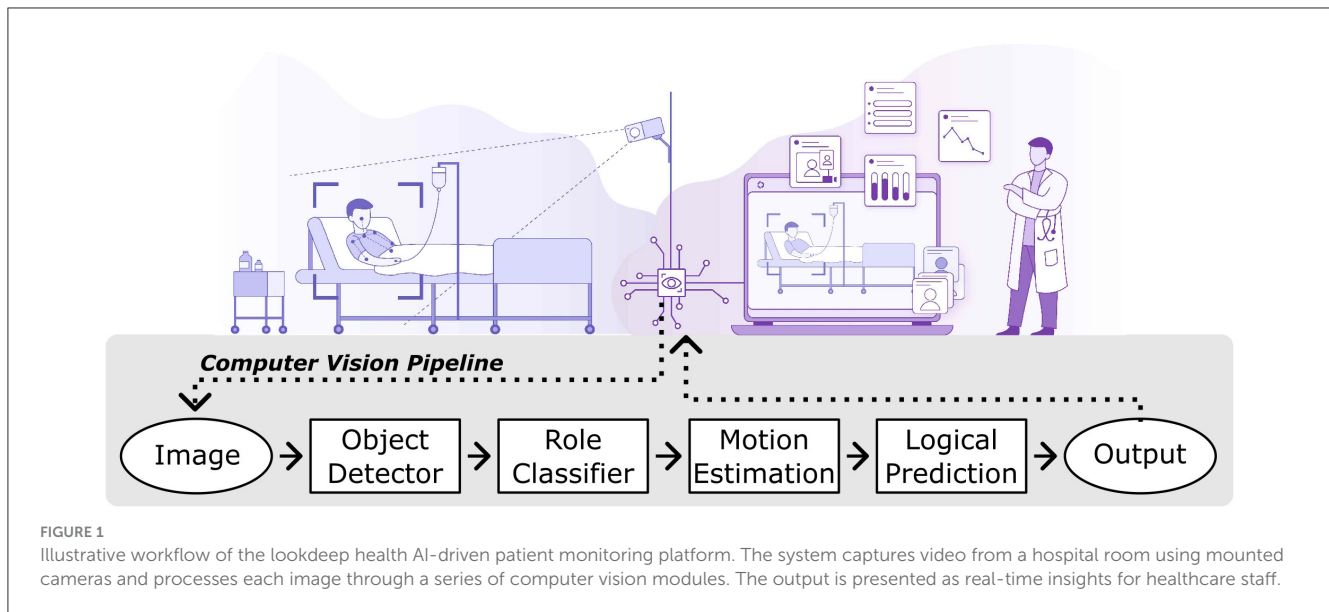
As shown in Figure 3, data collection spanned multiple years, with each subset contributing to the development and validation of the AI system, with some overlap between subsets.

2.1.2 Patient monitoring system overview

The LookDeep Health monitoring system processes video through a computer vision pipeline to detect, classify, and analyze key elements within the patient's room, providing actionable insights to healthcare staff (Figure 4). Key components include:

1. **Video data capture and preprocessing:** video data is captured at 1 frame per second (fps) by LookDeep Video Unit (LVU) devices deployed in patient rooms (Figure 5A). Data is preprocessed to reduce bandwidth and enable efficient analysis.
2. **Object detection and localization:** a custom-trained model detects key objects ("person", "bed", "chair") and localizes them with bounding boxes.
3. **Person-role classification:** detected "person" objects are further classified as "patient", "staff", or "other" using the same object detector model, by augmenting labels with role-specific information.
4. **Motion estimation:** dense optical flow estimates motion between consecutive frames, enabling activity tracking in specific regions (e.g. scene, bed, safety zone).
5. **Logical predictions:** high-level predictions (e.g. "person alone", "patient supervised by staff") are derived by applying rules to detection and motion data, with a 5-second smoothing filter to mitigate detection errors.

Inference results, including object detections, role classifications, motion estimation, and logical predictions, are securely stored in a Google cloud database for further analysis (e.g.

FIGURE 1
Illustrative workflow of the lookdeep health AI-driven patient monitoring platform. The system captures video from a hospital room using mounted cameras and processes each image through a series of computer vision modules. The output is presented as real-time insights for healthcare staff.

trend analysis). Anonymized frames are stored at regular intervals for quality assurance and model improvement.

2.1.3 Data anonymization

To ensure patient privacy in accordance with the Health Insurance Portability and Accountability Act (HIPAA) and institutional guidelines, all video data was processed to remove identifiable information. For training purposes, frames were face-blurred using a two-step procedure to maintain privacy while preserving relevant scene context:

1. **Manual labeling:** faces were manually labeled on fully-blurred images to create bounding boxes without exposing identifiable features.
2. **Local Gaussian blurring:** a strong Gaussian blur was applied to labeled facial regions, preserving scene context while anonymizing identities.

This approach was chosen to ensure privacy while balancing effective model training and validation. Additional obfuscation methods, such as pixelation or complete occlusion of faces, were considered but deemed not necessary for the intended use case. Data handling was conducted under a Business Associate Agreement (BAA) with participating hospitals.

2.2 Data collection

2.2.1 Video patient monitoring

LVU devices capture continuous video in RGB or near-infrared (NIR) mode, depending on ambient lighting. Each device is equipped with a CPU and Neural Processing Unit (NPU), capable of processing data at 1fps to minimize latency and reduce cloud processing requirements. Inference results are uploaded to a secured cloud database (Google BigQuery), with blurred frames stored separately for manual annotation. Camera placement varied based on room layout and clinical workflows (Figure 5B).

2.2.2 Annotations
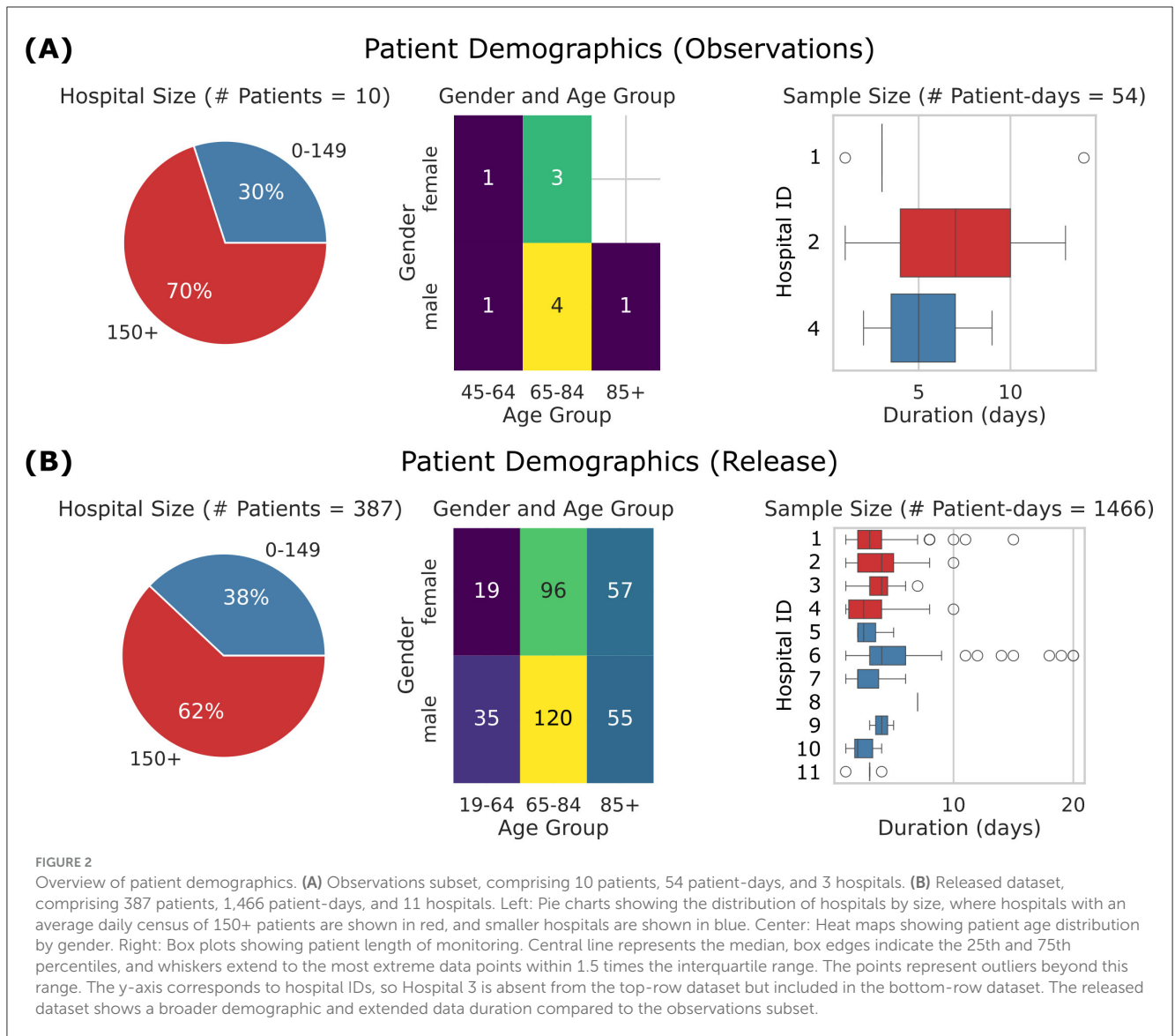#### 2.2.2.1 Frame-level labels

A professional labeling team manually annotated over 40,000 images with object bounding boxes, object properties, and scene-level tags (Figure 6). Objects were annotated with 2-d bounding boxes classed as "person", "bed", or "chair", and each "person" bounding box was also assigned a role of "patient", "staff", or "other". Scene level attributes were added for whether the patient was "in bed" or "not in bed", whether the camera was operating in IR mode, and whether the scene included "exception cases" in comparison to stated norms. Exception cases were applied in any instance of labeler uncertainty (e.g. difficult to see person, patient in street clothes, etc.); in instances of multiple exception cases being applicable, a single "frame exception" catch-all was used. Annotations and quality review were conducted using the Computer Vision Annotation Tool (CVAT, Corporation, 2023), and final QA was conducted using the FiftyOne tool (Moore and Corso, 2024).

2.2.2.2 Observation logs

Blurred video summaries for 10 patients (54 patient-days) were reviewed to log periods when the patient was alone. Logs included timestamps with 1-2 second precision (Figure 6), and underwent secondary quality assurance to provide feedback to labelers and fill out any missing periods.

2.3 Computer vision predictions

The LookDeep Health pipeline processes video data using custom-trained models to detect objects, classify person-role, and estimate motion at 1 fps. Preprocessing compresses frames to JPEG at 80% quality and resizes to a resolution of 1088x612 to reduce bandwidth consumption while still meeting downstream model requirements. Image processing is conducted using OpenCV (Bradski, 2000) and RKNN-toolkit (AI Rockchip, 2024).

FIGURE 2
Overview of patient demographics. **(A)** Observations subset, comprising 10 patients, 54 patient–days, and 3 hospitals. **(B)** Released dataset, comprising 387 patients, 1,466 patient–days, and 11 hospitals. Left: Pie charts showing the distribution of hospitals by size, where hospitals with an average daily census of 150+ patients are shown in red, and smaller hospitals are shown in blue. Center: Heat maps showing patient age distribution by gender. Right: Box plots showing patient length of monitoring. Central line represents the median, box edges indicate the 25th and 75th percentiles, and whiskers extend to the most extreme data points within 1.5 times the interquartile range. The points represent outliers beyond this range. The y-axis corresponds to hospital IDs, so Hospital 3 is absent from the top-row dataset but included in the bottom-row dataset. The released dataset shows a broader demographic and extended data duration compared to the observations subset.

1. **Object detection (person/bed/chair):** based on the YOLOv4 architecture (Bochkovskiy et al., 2020), the model identifies key objects in each frame, including "person", "bed", and "chair". Training models were initialized using COCO weights (Lin et al., 2014), then fine-tuned on labeled data. Input images were down-sampled to 608 × 608 with OpenCV's cubic interpolation method to fit model requirements. Since the models operate with a smaller fixed input size, increasing the resolution of input images would not significantly improve detection performance unless alternative patch-based approaches were considered. Additionally, the impact of input size on detection accuracy has been well-documented in the original YOLOv4 manuscript, which demonstrated stable performance across various input sizes. Training was conducted on NVIDIA 3070 GPU, and models were subsequently converted for execution on the Rockchip RKNN embedded in the LVU devices.

2. **Person classification (patient/staff/other):** during object detector training, bounding box labels were augmented to classify detected persons by role ("patient", "staff", "other").

Then, at inference time, each "person-" bounding box are re-labeled as "person", with the specific role saved in a separate classification field. Confidence scores for role classifications are derived by taking the highest detection confidence as the primary class and distributing residual scores across remaining classes to indicate potential alternate roles.

3. **Optical flow (motion estimation):** motion between frames was estimated using the Gunnar-Farneback dense optical flow algorithm, which calculates horizontal and vertical displacement for each pixel (Farnebäck, 2003). Optical flow inputs were converted to grayscale and down-sampled to 480x270 to ensure real-time execution. For each region of interest, average motion magnitude was calculated by averaging horizontal and vertical flow vectors, providing an indicator of activity intensity. This estimation does not require training and was implemented using OpenCV with fixed parameters: pyramid scale (pyr_scale = 0.5), number of pyramid levels (levels=3), window size (winsize = 15), number of iterations (iterations=3), size of pixel neighborhood used to find polynomial expansion (poly_n = 5), and the

FIGURE 3
Dataset overview and timeline of model updates. Progression of data collection and model updates for the LookDeep Health monitoring system. Single-frame analysis data collection spans a two year period—a broad temporal range for training and validation of object detection and classification tasks. Observation logging data, used for trend validation, was collected over a one year period. The publicly released dataset includes data from a more recent six month period, representing over 1,000 collective patient days. Model updates are indicated by numbered points.

standard deviation of the Gaussian used to smooth derivatives (poly_sigma = 1.2).

2.3.1 Additional components

2.3.1.1 Regions of interest (ROIs)

ROIs, such as "safety zones", provide contextual boundaries for monitoring. They are not predictive outputs themselves, but instead are used to track patient movements and boundary crossings. The "safety zone" was a polygonal region defined by the virtual monitor; its pixel mask is generated by expanding the boundary perimeter by 10% to ensure effective monitoring. Additional ROIs used by the system include the full scene and the detected bed.

2.3.1.2 Logical predictions

Logical predictions summarize patient status and interactions. These predictions were derived from a combination of object detection and role classification results and smoothed with a 5-second filter to mitigate intermittent detection errors.

- **Person alone:** *True* when the average number of detected people in the room is less than two.
- **Patient alone:** *True* when the average number of detected people in the room is less than two, and at least one person is classified as a patient.
- **Supervised by staff:** *True* when the average number of detected people in the room is two or more, and at least one person is classified as healthcare staff.

2.3.1.3 Trend predictions

Trends provide insights into immediate and long-term patient activity, aiding in risk identification and care planning. Hourly trends summarize patient behavior (e.g. "alone" or "moving") based

on aggregated logical predictions. For each one-hour interval, predictions were used to calculate the percentage of time the patient spent in key states like "alone," "supervised by staff," or "moving". These percentages were then plotted over time to visualize hourly trends in patient isolation or activity levels throughout the day. These trends provide a high-level overview of patient behavior, aiding in the identification of potential risks and informing care decisions.

2.3.1.3.1 "Assisted" trend predictions

A one-off analysis was conducted to simulate the system's performance when one of the predictions was known. The system's trend predictions based solely on AI inference were compared with those generated using a combination of AI inference and observation logs. For this comparison, "assisted" trends were created by integrating AI-predicted states for "moving" and "supervised by staff" with manually logged periods of "alone" status from the observation logs. This analysis was conducted across the multiple patients and hospitals included in the "Observation Logging" dataset.
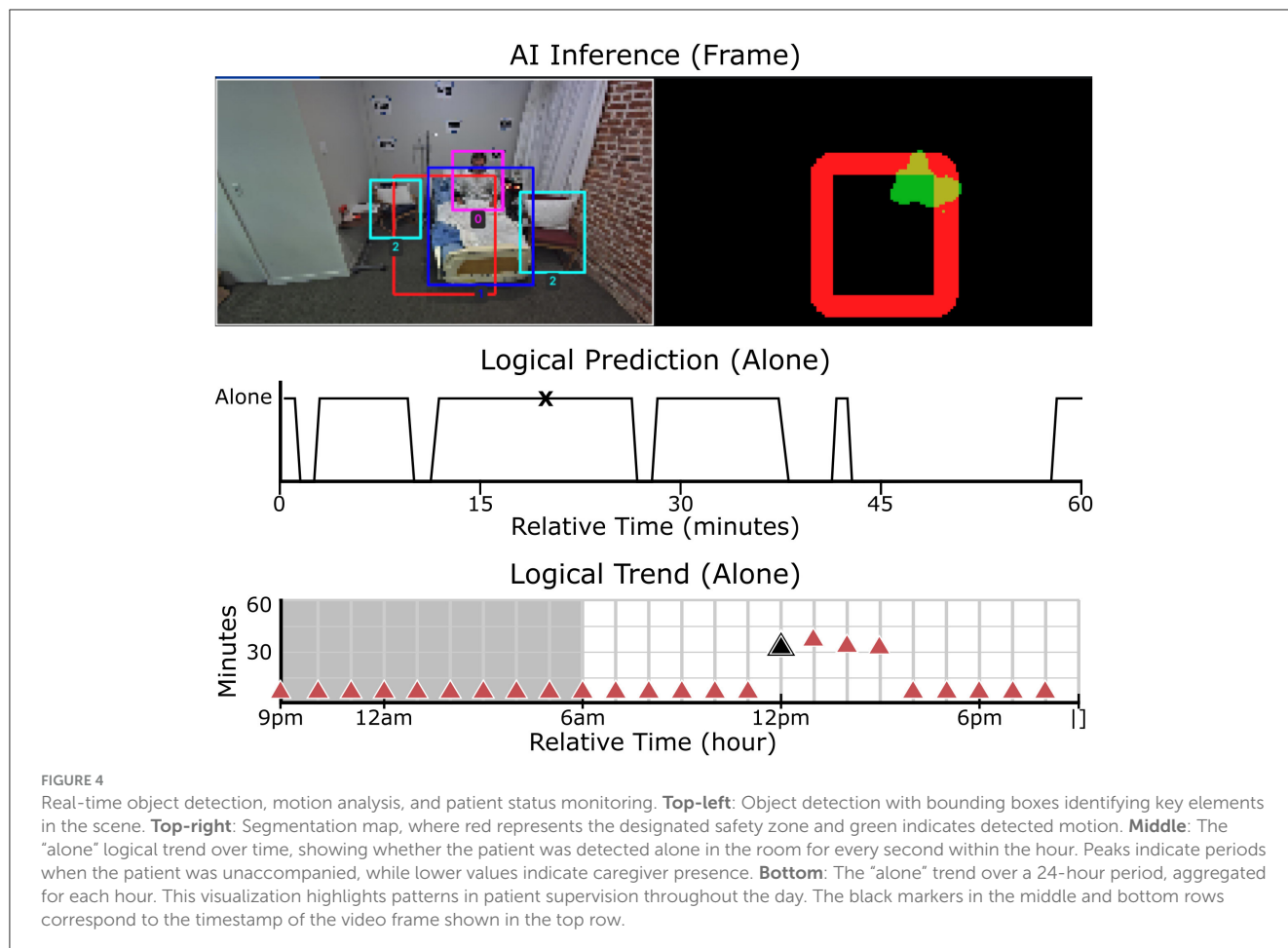
2.4 Evaluation

The performance of the AI-driven monitoring system was assessed through two primary methods: **image-level assessment** and **comparison against observation logs**. In the image-level assessment, each frame was analyzed against manual annotations to evaluate the accuracy of the system's object detection, person-role classification, and scene interpretation capabilities. In parallel, observation logs, created from anonymized video summaries of select patients, were compared against predicted trends to assess the system's ability to capture patient behavior patterns.

2.4.1 Frame-level analysis

Each model in the AI system was evaluated independently to assess its performance in object detection and classification tasks. Key performance metrics—precision, recall, and F1-score—were calculated to measure the accuracy and reliability of each model's predictions. Precision assessed the proportion of true positives among all predicted positives, recall measured the ability to identify all true positives, and the F1-score provided a balanced metric between precision and recall.

In addition to these direct object detection and classification tasks, the AI system also generated higher-level, "logical" predictions derived from these outputs. For example, the prediction "is patient alone" was inferred based on a combination of object detection results, such as the absence of healthcare staff within a defined proximity to the patient. These logical predictions were treated as classification tasks themselves, with their accuracy similarly evaluated using precision, recall, and F1-score metrics based on labeled image data. This multi-layered approach allowed us to thoroughly validate both the core object detection functions of each model and the system's ability to interpret and apply these outputs to patient monitoring tasks.

FIGURE 4
Real-time object detection, motion analysis, and patient status monitoring. **Top-left**: Object detection with bounding boxes identifying key elements in the scene. **Top-right**: Segmentation map, where red represents the designated safety zone and green indicates detected motion. **Middle**: The "alone" logical trend over time, showing whether the patient was detected alone in the room for every second within the hour. Peaks indicate periods when the patient was unaccompanied, while lower values indicate caregiver presence. **Bottom**: The "alone" trend over a 24-hour period, aggregated for each hour. This visualization highlights patterns in patient supervision throughout the day. The black markers in the middle and bottom rows correspond to the timestamp of the video frame shown in the top row.

2.4.2 Trend analysis

Trend analysis was conducted by comparing the system's inference-derived metrics to ground truth metrics recorded in observation logs, with both datasets aggregated by patient-day. Unlike the hourly trends shown in Figure 4, analysis was conducted at the per-second level to ensure accurate alignment between AI predictions and observation logs. The primary metric for this analysis was logistic regression accuracy, which assessed the AI system's ability to predict observed behaviors within three time periods: daytime (6 am to 9 pm), nighttime (9 pm–6 am), and the full 24-hour period. In cases where only a single class (e.g. "alone" or "not alone") was present within a specific time period, logistic regression was not feasible. Instead, a manual accuracy score was computed, to allow for consistent accuracy measurements across all time intervals. This score is defined as the proportion of matching values between the AI predictions and ground truth.
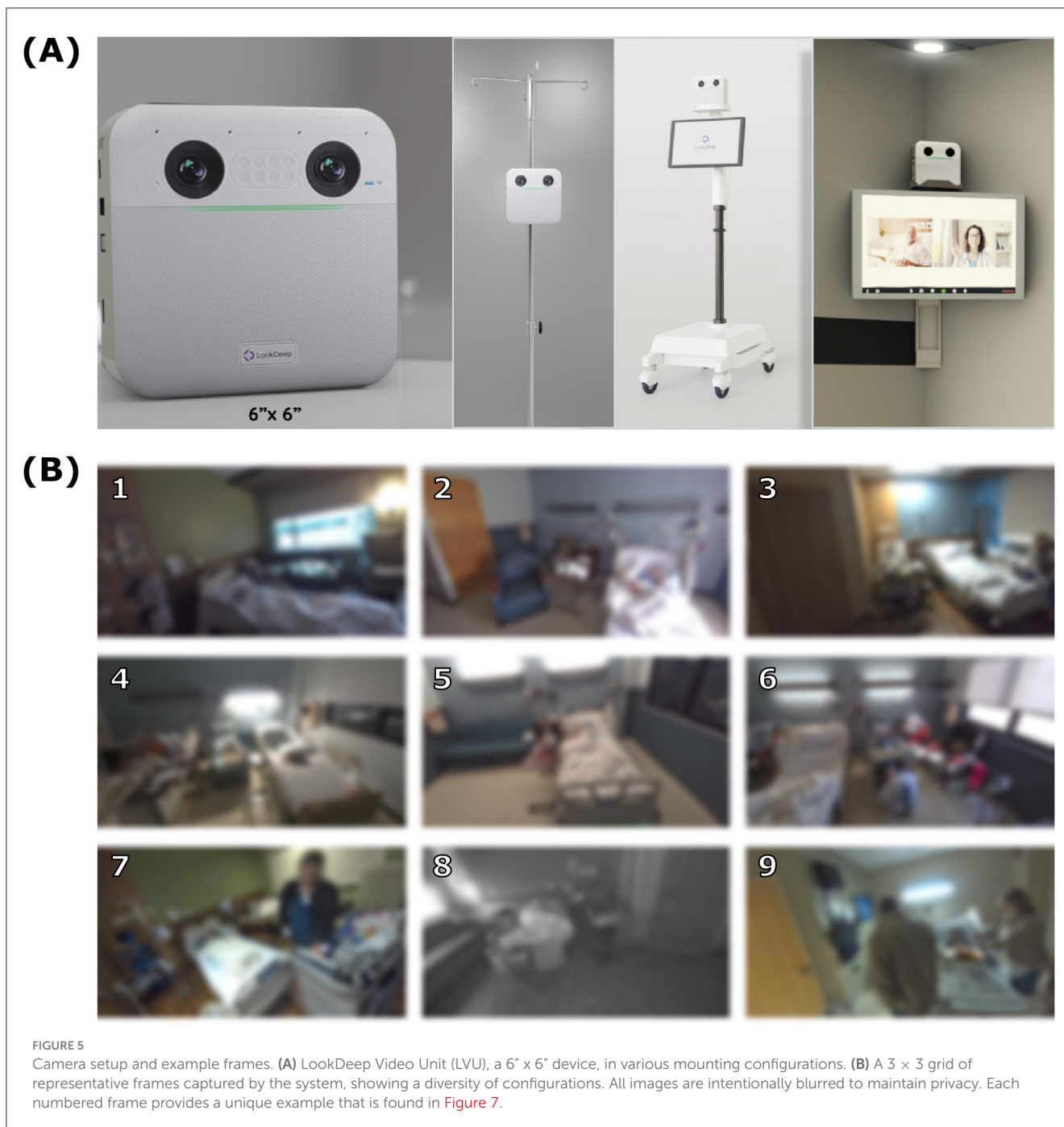
Focusing on the "alone" binary behavior trend enables an assessment of the alignment between AI predictions and real-world observations. This analysis validated the AI system's effectiveness in capturing hourly patient behavior trends, underscoring its potential utility in real-time patient monitoring and early detection of deviations from expected patterns.

2.4.3 Camera position meta-analysis

Since cameras were mounted on mobile carts rather than fixed positions, there was variability in camera setup across patients and hospital rooms (Figure 4B). To explore the potential impact of this variability, labeled bed locations were used to estimate each camera's position relative to the hospital bed. Distributions of the labeled bed area and size within each frame, along with the centroid location of the bed relative to the camera's field of view are plotted in Figure 7. These distributions provide an indirect measure of camera position.

This exploratory analysis helped identify patterns and variations in camera setups across different monitoring sessions. However, this information was observational and used only to understand positional variability; no specific adjustments were made during model training or evaluation to account for different camera positions. The results underscore the robustness of our models in handling diverse camera perspectives, as the system maintained consistent detection performance despite these variations.

FIGURE 5
Camera setup and example frames. **(A)** LookDeep Video Unit (LVU), a 6″ x 6″ device, in various mounting configurations. **(B)** A 3 × 3 grid of representative frames captured by the system, showing a diversity of configurations. All images are intentionally blurred to maintain privacy. Each numbered frame provides a unique example that is found in Figure 7.

3 Results

3.1 Frame-level analysis

3.1.1 Object detection, role identification, and patient isolation classification

The evaluations demonstrated that the custom-trained computer vision models perform robustly in real-world hospital settings, achieving high precision across key object detection and classification tasks. We compared five production models alongside a baseline model using an off-the-shelf YOLOv4 configuration (Table 1). Each production model corresponds to

a different release, with progressively larger and more refined training datasets incorporated over time (Figure 3). This iterative refinement led to increased model accuracy and adaptability in real-world hospital settings. To ensure consistency, all frame-level analysis was conducted on 10,000 frames collected over a two year period. This representative sample, excluded from model training and validation, highlights the incremental improvements achieved by expanding training datasets across model versions.
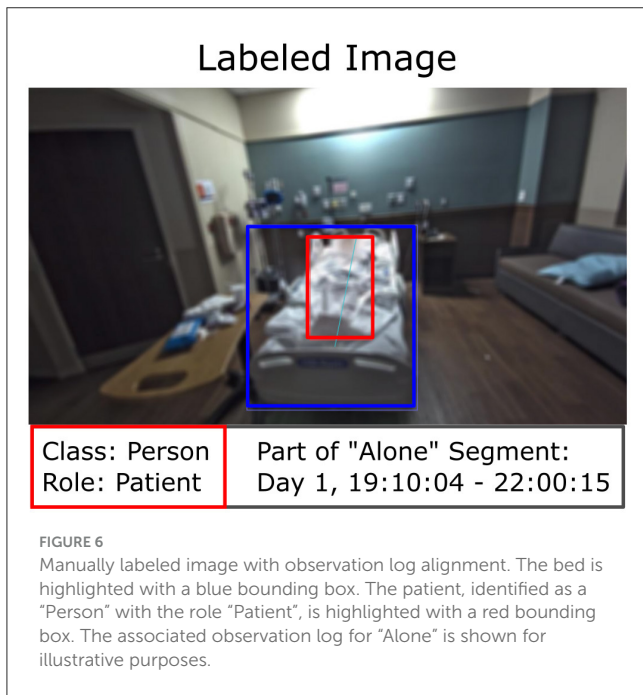
As newer models were released, the training set was expanded to include additional annotated data, allowing each successive model to capture more complex and diverse scenarios encountered in hospital environments. The most recent fine-tuned model (v5)

achieved an **F1-score of 0.91** for detecting "person", notably surpassing the baseline YOLOv4 model score of 0.41 (Table 2). Across all object classes–including beds, furniture, and other room elements–the v5 model demonstrated an **F1-score of 0.92**, reflecting a high degree of accuracy and consistency across diverse object types.

In addition to object detection, the system was evaluated on a three-class person-role classification task, distinguishing between patients, healthcare staff, and visitors within the camera's field of view. The v5 model demonstrated particularly strong performance for the "patient" class, achieving an **F1-score of 0.98**, which reflects

its high accuracy in identifying patients specifically (Table 2). Accurate person-role classification is essential for monitoring patient interactions and ensuring appropriate caregiving behaviors, as it enables the system to capture not only the presence of individuals but also their roles. Focusing on the "patient" class, the high F1-score underscores the model's robustness in tracking patient activity and interactions, which are critical for effective continuous monitoring in dynamic hospital environments.

The downstream classification task of identifying whether a patient was "alone" in the room showed similarly strong results, with the v5 model achieving an **F1-score of 0.92** (Table 2). This classification task, essential for monitoring patient isolation, consistently improved with each new production release, as more comprehensive training data contributed to better model accuracy. These results confirm the advantage of iterative model refinement

FIGURE 6
Manually labeled image with observation log alignment. The bed is highlighted with a blue bounding box. The patient, identified as a "Person" with the role "Patient", is highlighted with a red bounding box. The associated observation log for "Alone" is shown for illustrative purposes.

TABLE 1 Performance metrics of successive model versions for object detection.

Model version	Fine-tuning data size	Object detection (all)	
		Precision	F1
YOLOv4 (baseline)	n/a	0.84	0.59
Model v1 (2022 Q1)	+700	0.97	0.74
Model v2 (2023 Q2)	+2,474	0.98	0.83
Model v3 (2023 Q3)	+10,133	0.97	0.83
Model v4 (2024 Q1)	+28,914	0.98	0.91
Model v5 (2024 Q2)	+34,239	0.97	0.92

Summary of precision and F1-scores across different versions of the LookDeep Health AI model, highlighting improvements in key tasks as the training data increased. The baseline YOLOv4 model demonstrates initial performance levels, while successive versions (Models v1 to v5) show incremental gains in object detection. With each model iteration, higher precision and F1-scores indicate enhanced detection accuracy and classification reliability, underscoring the impact of additional data and model refinement on real-time patient monitoring capabilities. Evaluation was performed on a fixed dataset containing 10k images.
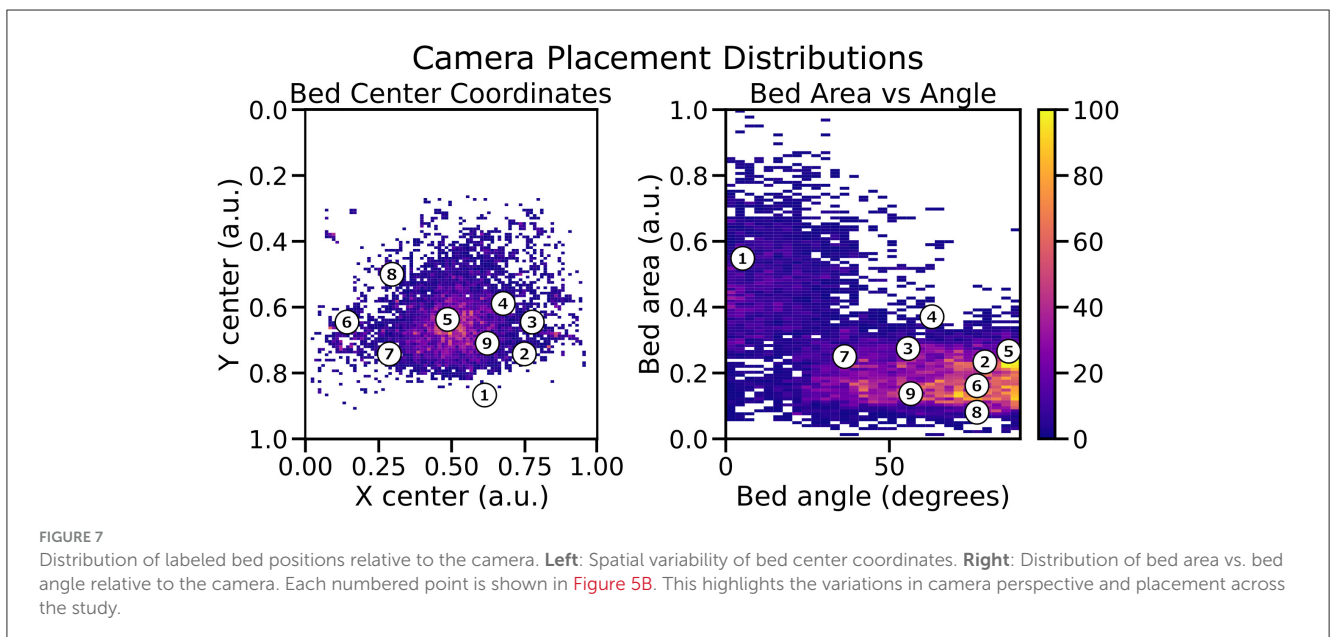

FIGURE 7
Distribution of labeled bed positions relative to the camera. **Left**: Spatial variability of bed center coordinates. **Right**: Distribution of bed area vs. bed angle relative to the camera. Each numbered point is shown in Figure 5B. This highlights the variations in camera perspective and placement across the study.

TABLE 2 Performance metrics of successive model versions for object detection (person), role classification, and "patient alone" classification.

Model version	Object detection (person)		Role classification (patient F1)	"Patient alone" classification (F1)
	Precision	F1		
YOLOv4 (baseline)	0.98	0.41	n/a	0.28
Model v1 (2022 Q1)	0.98	0.85	n/a	0.86
Model v2 (2023 Q2)	0.97	0.89	n/a	0.91
Model v3 (2023 Q3)	0.97	0.86	0.97	0.88
Model v4 (2024 Q1)	0.97	0.91	0.98	0.94
Model v5 (2024 Q2)	0.96	0.91	0.98	0.92

Additional results corresponding to Table 1 are presented here, focusing on object detection of persons, role classification, and "patient alone" classification tasks.

TABLE 3 Performance comparison of models on unblurred vs. face-blurred images across versions.

Model version	Evaluation data size	Unblurred images (F1)	Face-blurred images (F1)	Δ F1
Model v3 (2023 Q3)	2,135	0.81	0.85	+0.04
Model v4 (2024 Q1)	1,809	0.86	0.90	+0.04
Model v5 (2024 Q2)	1,226	0.89	0.91	+0.02

Evaluation of model performance on unblurred and face-blurred images across different versions. The F1-score measures the model's performance, with the "Δ F1" column showing the gap between unblurred and face-blurred images. A Δ value closer to 0 indicates better consistency in model performance between unblurred and face-blurred images.

and dataset expansion, with each production release yielding models that are better adapted to the variability and demands of real-world hospital settings.

3.1.2 Impact of privacy-preserving blurring on model consistency

The performance consistency of the models across unblurred and face-blurred images was evaluated using the Δ metric, which represents the F1-score difference between the two image types (Table 3). Across all model versions, the Δ values were relatively small, indicating that face-blurring–a common privacy-preserving preprocessing step–had minimal impact on model accuracy. For versions v3 and v4, the Δ value was +0.04, while in v5 it decreased to +0.02, suggesting improved robustness to blurring as the training data volume increased.

A smaller Δ value is desirable as it indicates that the model performs consistently regardless of whether the images are unblurred or face-blurred. The reduction in Δ for v5 highlights the value of larger, more diverse training datasets in ensuring that the models generalize well across different image types. This is particularly important in hospital settings, where preserving patient privacy often necessitates the use of face-blurred images. The ability to maintain high accuracy in such scenarios ensures the system's practicality and reliability for real-world deployment.

These results demonstrate that the models not only achieve high accuracy but also exhibit resilience to variations introduced by privacy-preserving preprocessing, a key requirement for scalable applications in healthcare environments.

3.1.3 Object detector performance by IR mode

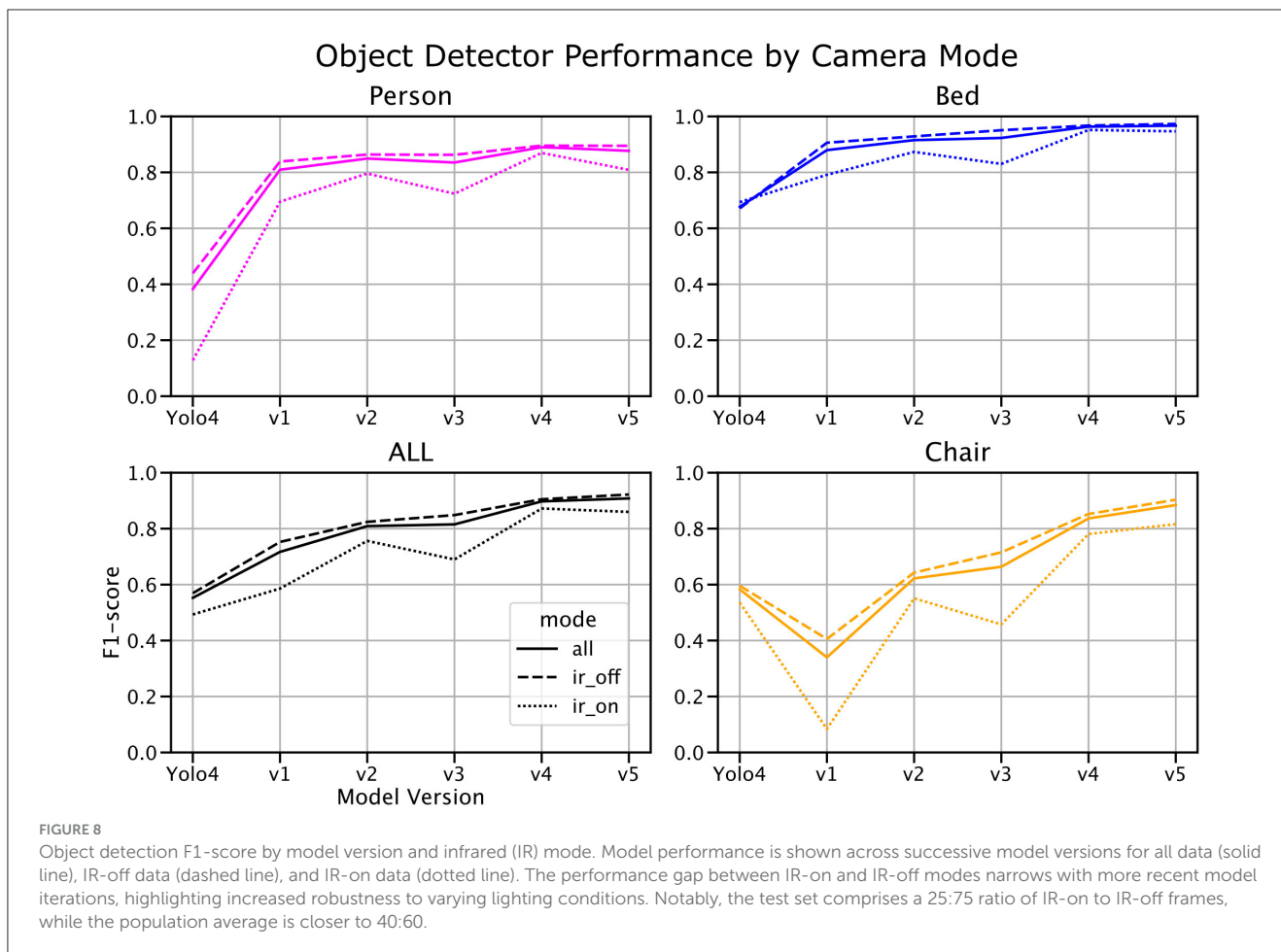We analyzed the impact of IR mode on object detection performance by comparing F1-scores across different model versions, broken down into all data, IR-on data, and IR-off data (Figure 8). Results demonstrate a clear trend of increasing F1-scores with newer model versions across all conditions. Notably, the performance gap between IR-on and IR-off scenarios decreases with successive model iterations, indicating improved model robustness to variations in lighting conditions.

At baseline, object detection performance in IR-on scenarios lagged significantly behind IR-off scenarios. However, with the latest model version, this gap narrowed substantially, suggesting that additional training data and model refinements have enhanced the system's ability to generalize across lighting conditions. Despite these improvements, it is worth noting that the test set contains an approximate 25:75 ratio of IR-on to IR-off frames, whereas the population average is closer to 40:60. This imbalance may partially account for residual performance differences and highlights the need for more balanced representation in future datasets.

These findings underscore the importance of accounting for lighting variability in real-world hospital environments and demonstrate the system's potential to adapt to challenging conditions such as low-light monitoring.

3.2 Trend analysis

Inference-derived trends for the "patient alone" metric were compared against observation logs to evaluate the system's ability to accurately capture real-world patterns (Figure 9). This trend analysis utilized data from earlier stages of the project when base models with lower performance were deployed. Specifically, the object detectors used for these inferences had an F1-score of 0.85 for "person" detection, which is lower than the performance of the latest models. Despite this, the analysis showed strong

FIGURE 8
Object detection F1-score by model version and infrared (IR) mode. Model performance is shown across successive model versions for all data (solid line), IR-off data (dashed line), and IR-on data (dotted line). The performance gap between IR-on and IR-off modes narrows with more recent model iterations, highlighting increased robustness to varying lighting conditions. Notably, the test set comprises a 25:75 ratio of IR-on to IR-off frames, while the population average is closer to 40:60.

alignment with ground truth data, achieving an **average logistic regression/manual accuracy of 0.84 ± 0.13 during daytime, 0.80 ± 0.16 at nighttime, and 0.82 ± 0.15 across all times**. These results highlight the robustness of the AI system in capturing patient isolation trends, even when using earlier model versions with lower baseline performance.

This accuracy indicates that, even with slightly reduced detection precision in the older models, the system could reliably capture general patterns in patient isolation behavior. The standard deviation (± 0.15) reflects some variability in accuracy across different times of day and patient conditions, possibly influenced by factors such as changing camera angles or environmental conditions. As shown in the normative hourly trends (Figure 10), discrepancies between labeled and AI-inferred "alone" data are more pronounced during nighttime hours, but these differences have minimal impact on the broader trend patterns. For both "Alone and Moving" and "Supervised by Staff" metrics, the AI inferences closely align with label-assisted data, amounting to an **average error of 1–2 min per hour**. This consistency underscores the model's robustness in capturing meaningful patient-alone trends and suggests that any nighttime performance gaps in the "alone" inference do not significantly compromise the overall accuracy. These results highlight the model's potential for improved trend detection as newer, refined models are applied to subsequent data.

4 Discussion

4.1 Implications for clinical practice

The findings of this study underscore the potential for AI-enabled patient monitoring systems to enhance clinical practice through continuous, real-time monitoring. Traditional in-person observations are limited by the time constraints of healthcare staff, who spend limited hours directly interacting with each patient. By providing continuous monitoring, the LookDeep Health platform enables staff to detect patterns that would otherwise go unnoticed, such as extended periods of patient isolation, movement patterns that might indicate a risk of falls, pressure injuries, or irregular interactions with staff. Real-time alerts based on these observations could prompt timely interventions, potentially improving patient safety and outcomes.

Moreover, the data collected by this system can inform trend analysis on a population level, supporting hospital resource allocation and staffing decisions. For instance, identifying times of day when patients are frequently unsupervised could guide adjustments in staffing or the deployment of additional monitoring resources to high-risk patients. Beyond staffing, the system's insights into patient mobility patterns–such as time spent in bed, in a chair, or walking around the room–can help identify markers of successful recovery and readiness for discharge, contributing

FIGURE 9

Comparison of average minutes alone per hour across all patients, a typical patient, and an atypical patient. The average minutes patients spent alone per hour, comparing ground truth (black squares) and AI inference (red triangles) across three scenarios: all patients **(top)**, a typical patient **(middle)**, and an atypical patient **(bottom)**. The x-axis shows the hour of the day, while the y-axis indicates the average minutes alone per hour. The shaded region represents nighttime hours (9 pm−6 am). For all patients, AI inference closely aligns with ground truth during the day but shows less accuracy at night. An example of a typical and an atypical patient is shown to illustrate the variability in alone time patterns across individual patients. Unlike average trends, this atypical patient exhibits an overprediction of alone time at night, highlighting the need for further model refinement to capture individual patient behaviors accurately.

FIGURE 10

Average trends for all observed patients. Hourly trends are compared across two metrics: alone and moving **(top)** and supervised by staff **(bottom)**. Fully AI-inferred data ("Inference Only") is plotted with red triangles. "Assisted" data ("+Labels") is plotted with blue circles—for these data, ground truth "alone" status was used instead of AI-inference. The x-axis indicates the hour of the day, while the y-axis shows the average minutes per hour. The shaded region represents nighttime hours (9 pm−6 am). Although there is a discrepancy between labeled and inferred data for Alone, particularly during nighttime hours, the downstream impact on overall trend accuracy appears minimal (1−2 min per hour).

to improved patient outcomes. These mobility insights could also support the development of best practices for post-procedure mobility, tailored to specific surgeries or treatments, to enhance patient recovery. Altogether, these data-driven insights promote a more efficient, personalized approach to patient care, potentially improving patient satisfaction and clinical outcomes.

4.2 Impact of face-blurring on model performance

While the evaluation of model performance on both unblurred and face-blurred images provides valuable insights, it is important to note that face-blurring is applied only during training and
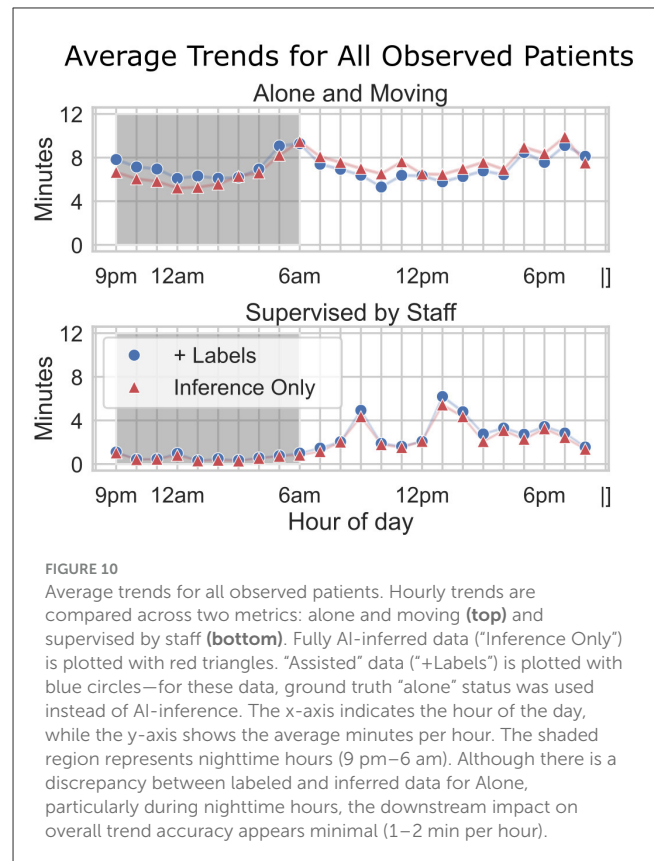
evaluation phases. In real-world deployment, the model will encounter unblurred images as it monitors patients in hospital settings, making this distinction critical to understanding its practical performance. The small Δ values observed across different model versions indicate that the models have been designed to handle face-blurred images without significant degradation in performance. The reduced Δ in the latest version (v5), attributed to increased training data volume, demonstrates improved resilience to face-blurring. However, further studies are needed to assess the model's performance in unblurred scenarios, particularly in environments where face-blurring images for training and evaluation is not an option. This approach ensures privacy during development while maintaining practical deployment fidelity, as real-time monitoring operates on unblurred frames.

4.3 Variation in camera setup

The LookDeep Health patient monitoring platform was deployed in real-world hospital settings with cameras mounted on mobile carts rather than fixed positions, resulting in variation in camera angles, distances, and perspectives across different patient rooms. This variability introduced potential challenges in maintaining consistent object detection and classification accuracy, as model performance can be influenced by changes in camera field of view and positioning relative to the bed. To mitigate these effects, we conducted a camera position meta-analysis using metadata on labeled bed area and centroid location to estimate the approximate

camera placement within each room. Our analysis confirmed that, despite positional differences, the model consistently achieved reliable performance across object detection and classification tasks, demonstrating its robustness to spatial variability. However, this setup presents limitations in controlling for optimal camera positioning, a factor that future studies with standardized camera setups could explore further to minimize variability and enhance model reliability.

4.4 Nuanced differences in time coverage of analyses

A key aspect of this study is the variation in time coverage across different datasets, reflecting the evolving nature of data collection and model validation in real-world hospital settings. The observation logs dataset, which provided ground truth for logical trend validation, was collected exclusively in 2023. In contrast, frame-level annotations for evaluating object detection and person-role classification were gathered over a more extended period from 2022 to 2024. Additionally, the publicly released dataset comprises data collected from a 6 month span across 2024, representing over 1,000 collective patient days across multiple hospitals.

These differences in collection periods introduce nuances in interpretation. For instance, frame-level evaluations benefit from the broader time span, capturing a variety of hospital conditions and patient behaviors across seasons and changing workflows. However, trend analyses were constrained to the observation log time frame, which may limit the ability to generalize trends across the entire study period. Similarly, the released dataset reflects data from the latter phase of the study, aligning with the most refined models but excluding early-stage model iterations.

These variations in time coverage highlight the need to contextualize each analysis within its specific time frame. Future studies could benefit from aligning data collection periods across all evaluation methods, ensuring that models validated on frame-level tasks are continuously validated against trend and behavioral analyses for consistent performance insights over time.

4.5 Challenges and limitations

Several challenges and limitations were encountered in this study. First, the variability in camera setup, as mentioned earlier, introduces potential inconsistencies in model performance due to changing perspectives and distances. While our metadata analysis mitigated this to some extent, a standardized camera setup would likely yield more consistent results.

Second, while the LookDeep Health system demonstrated strong performance in object detection and role classification, real-time video processing presents computational challenges that require balancing accuracy and processing speed. Our use of onboard CPU and NPU on LVU devices provided sufficient processing capabilities for 1 fps inference; however, the scalability of

such a setup may be constrained in larger hospital systems requiring higher frame rates for finer details.

Third, the dataset collected in this study primarily consists of high-risk fall patients, which may limit the generalizability of findings to broader patient populations - for example, high-risk patients exhibit limited mobility compared to other patient groups. Additionally, the analysis was conducted on older model versions for some trend analyses, potentially lowering the accuracy of trend detection. Although model refinements are expected to improve results, these differences in model versions should be considered when interpreting the findings.

Lastly, maintaining patient privacy is paramount in continuous video monitoring systems. While the LookDeep Health platform anonymizes all video and stores de-identified data, ongoing attention to data privacy and compliance with healthcare regulations is essential for future deployments in clinical environments.

4.6 Suggestions for future research

While this study provides a foundation for understanding the impact of AI-driven patient monitoring, further research is warranted to explore additional facets of this technology. Future studies could investigate:

- **Enhanced edge case handling:** expanding training datasets to include more examples of diverse scenarios, such as low-light conditions and atypical patient behaviors, could improve model robustness in challenging environments.
- **Advanced deep learning techniques:** integrating more sophisticated deep learning architectures like transformer-based architectures or temporal models could enhance the detection of subtle anomalies, while adaptive pipelines could improve real-time robustness in dynamic hospital environments.
- **Refining architectures and guardrails:** future work could involve refining architectures to detect edge cases more accurately, tracking patterns in prediction errors, and incorporating confidence-based guardrails to prevent catastrophic failures. Such guardrails could include alerts when model confidence is unexpectedly low for consecutive predictions.
- **Higher frame rates and computational scaling:** evaluating the potential for higher frame rates or adaptive frame rate technology to improve real-time responsiveness, particularly in high-activity environments.
- **Standardization of camera placement:** testing standardized, fixed camera setups across patient rooms aims to minimize positional variability and improve model consistency. Although standardization can reduce variability, embracing the inherent diversity of setups may enhance model robustness for real-world applications.
- **Expanded patient cohorts:** extending the analysis to include a wider range of patient demographics and conditions to assess generalizability and adapt the system to diverse populations.
- **Interoperability with hospital systems:** future iterations of the system could integrate more seamlessly with hospital

workflows by automating real-time alerts that directly sync with electronic health record (EHR) systems. For example, patient-specific alerts could be tagged to relevant EHR fields, enabling clinicians to view contextual video data alongside medical records. Additionally, the system could support interoperability with existing hospital tools, such as nurse call systems, to streamline the clinical response to high-risk situations.

These research directions, alongside continued refinement of computer vision models and monitoring systems, will be essential for advancing the practical application of AI in patient monitoring and driving further improvements in healthcare delivery.

5 Conclusion

AI integration in medical imaging is advancing personalized patient treatment but still faces challenges related to effectiveness and scalability. This work demonstrates the potential of computer vision as a foundational technology for continuous and passive patient monitoring in real-world hospital environments.

The contributions of this study are two-fold. First, we introduce the LookDeep Health patient monitoring platform, which leverages computer vision models to monitor patients continuously throughout their hospital stay. This platform scales to support a large number of patients and is designed to handle the complexities of hospital-based data collection. Using this system, we have compiled a unique dataset of computer vision predictions from over 300 high-risk fall patients, spanning 1,000 collective days of monitoring. To encourage further exploration in the field, we released this anonymized dataset publicly at https://github.com/lookdeep/ai-norms-2024.

Second, we rigorously validated the AI system, demonstrating strong performance in image-level object detection and person-role classification tasks. Our analysis also confirms a positive alignment between inference-derived trends and human-observed behaviors on a patient-hour basis, underscoring the reliability of the AI system in capturing patient activity trends. This evaluation can serve as a benchmark for future studies, providing a standard set of criteria for assessing the performance of AI-driven patient monitoring systems.

The extensive dataset and rigorous validation of the LookDeep Health platform highlight the feasibility and impact of continuous patient monitoring through video. By offering real-time insights into patient activity and isolation patterns, continuous monitoring has the potential to reduce fall risks by alerting staff to high-risk situations as they unfold. Beyond improving patient safety, these insights support more efficient staffing and resource allocation, allowing hospitals to adjust care based on real-time patient needs. This predictive capability also aids administrators in managing bed occupancy and optimizing patient flow, particularly during peak times, thus enhancing the responsiveness, efficiency, and scalability of the healthcare system. This work paves the way for future advancements in AI-driven healthcare solutions, promising scalable, data-informed insights to elevate patient care and hospital management.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

The data used in this retrospective study was collected from patients admitted to one of eleven hospital partners across three different states in the USA. The study and handling of data followed the guidelines provided by CHAI standards. Access to this data was granted to the researchers through a Business Associate Agreement (BAA) specifically for monitoring patients at high risk of falls. In compliance with the Health Insurance Portability and Accountability Act (HIPAA), patients provided written informed consent for monitoring as part of their standard inpatient care. To ensure patient privacy, all video data was blurred prior to storage, and no identifiable information is included in this work. Face-blurred frames were used only for training purposes. Faces were manually labeled on fully-blurred images, and the raw images were then treated with a local Gaussian blur in the facial regions, ensuring privacy without compromising model training quality. The outcomes of this analysis did not influence patient care or clinical outcomes.

Author contributions

Funding

Acknowledgments

Conflict of interest

PG, PR, TT, TW, MC, and NS are current or former employees of LookDeep Health, the company that provided the tools used in this study. LookDeep Health was involved in data collection and analysis and reviewed the final manuscript prior to submission. The authors declare no other conflicts of interest related to this work.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. We acknowledge the assistance of OpenAI's ChatGPT (version 4, model GPT-4-turbo) in refining the text of this manuscript. This generative AI technology was accessed through OpenAI's platform and used to improve clarity and organization in the presentation of the research.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abbe, J. R., and O'Keeffe, C. (2021). Continuous video monitoring: implementation strategies for safe patient care and identified best practices. *J. Nurs. Care Qual.* 36, 137–142. doi: 10.1097/NCQ.0000000000000502

AI Rockchip (2024). *Fuzhou Rockchip Electronics Co. RKNN-Toolkit.* Available online at: https://github.com/rockchip-linux/rknn-toolkit (accessed May 26, 2022).

Avogaro, A., Cunico, F., Rosenhahn, B., and Setti, F. (2023). Markerless human pose estimation for biomedical applications: a survey. *Front. Comp. Sci.* 5:1153160. doi: 10.3389/fcomp.2023.1153160

Bajwa, J., Munir, U., Nori, A., and Williams, B. (2021). Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc. J.* 8, e188–e194. doi: 10.7861/fhj.2021-0095

Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv [Preprint].* arXiv:2004.10934.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's J. Softw. Tools.* 120, 122–125.

Capo-Lugo, C. E., Young, D. L., Farley, H., Aquino, C., McLaughlin, K., Colantuoni, E., et al. (2023). Revealing the tension: The relationship between high fall risk categorization and low patient mobility. *J. Am. Geriatr. Soc.* 71, 1536–1546. doi: 10.1111/jgs.18221

Chae, W., Choi, D.-W., Park, E.-C., and Jang, S.-I. (2021). Improved inpatient care through greater patient-doctor contact under the hospitalist management approach: a real-time assessment. *Int. J. Environ. Res. Public Health* 18:5718. doi: 10.3390/ijerph18115718

Chen, K., Gabriel, P., Alasfour, A., Gong, C., Doyle, W. K., Devinsky, O., et al. (2018). Patient-specific pose estimation in clinical environments. *IEEE J. Transl. Eng. Health Med.* 6, 1–11. doi: 10.1109/JTEHM.2018.2875464

Corporation, C. (2023). *Computer Vision Annotation Tool (cvat).* Zenodo. Available online at: https://zenodo.org/records/8416684

Davenport, T., and Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthc. J.* 6, 94–98. doi: 10.7861/futurehosp.6-2-94

Davoudi, A., Malhotra, K. R., Shickel, B., Siegel, S., Williams, S., Ruppert, M., et al. (2019). Intelligent icu for autonomous patient monitoring using pervasive sensing and deep learning. *Sci. Rep.* 9:8020. doi: 10.1038/s41598-019-44004-w

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., et al. (2021). Deep learning-enabled medical computer vision. *NPJ Digit. Med.* 4:5. doi: 10.1038/s41746-020-00376-2

Farnebäck, G. (2003). "Two-frame motion estimation based on polynomial expansion," in *Image Analysis: 13th Scandinavian Conference, SCIA* (Halmstad: Springer), 363–370.

Haque, A., Milstein, A., and Fei-Fei, L. (2020). Illuminating the dark spaces of healthcare with ambient intelligence. *Nature* 585, 193–202. doi: 10.1038/s41586-020-2669-y

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). "Microsoft coco: Common objects in context," in *Computer Vision-ECCV 2014: 13th European Conference* (Zurich: Springer), 740–755.

Lindroth, H., Nalaie, K., Raghu, R., Ayala, I. N., Busch, C., Bhattacharyya, A., et al. (2024). Applied artificial intelligence in healthcare: a review of computer vision technology application in hospital settings. *J. Imag.* 10:81. doi: 10.3390/jimaging10040081

Mascagni, P., Alapatt, D., Sestini, L., Altieri, M. S., Madani, A., Watanabe, Y., et al. (2022). Computer vision in surgery: from potential to clinical value. *NPJ Digit. Med.* 5:163. doi: 10.1038/s41746-022-00707-5

Moore, B. E., and Corso, J. J. (2024). *Fiftyone.* Available online at: https://www.voxel51.com/fiftyone/ and https://github.com/voxel51/fiftyone (accessed April 20, 2022).

Parker, S., Gilstrap, D., Bedoya, A., Lee, P., Deshpande, K., Gabriel, P., et al. (2022). "Continuous artificial intelligence video monitoring of icu patient activity for detecting sedation, delirium and agitation," in *C35. Topics in Critical Care and Respiratory Failure* (Washington DC: American Thoracic Society), A5719–A5719.

Posch, C., Serrano-Gotarredona, T., Linares-Barranco, B., and Delbruck, T. (2014). Retinomorphic event-based vision sensors: bioinspired cameras with spiking output. *Proc. IEEE* 102, 1470–1484. doi: 10.1109/JPROC.2014.2346153

Watzlaf, V. J., Moeini, S., and Firouzan, P. (2010). Voip for telerehabilitation: A risk analysis for privacy, security, and hipaa compliance. *Int. J. Telerehabil.* 2:3. doi: 10.5195/ijt.2010.6056

Westbrook, J. I., Duffield, C., Li, L., and Creswick, N. J. (2011). How much time do nurses have for patients? a longitudinal study quantifying hospital nurses' patterns of task time distribution and interactions with health professionals. *BMC Health Serv. Res.* 11, 1–12. doi: 10.1186/1472-6963-11-319

Wilson, J. E., Mart, M. F., Cunningham, C., Shehabi, Y., Girard, T. D., MacLullich, A. M., et al. (2020). Delirium. *Nat. Rev. Dis. Prim.* 6:90. doi: 10.1038/s41572-020-00223-4