

## PAPER

# Development of a Mobile Video Platform for Diagnosing Human Motor Anomalies

Mykyta Kyt ,  
Valentyn Yesilevskyi 

Kharkiv National University  
of Radio Electronics,  
Kharkiv, Ukraine

[mykyta.kyt@nure.ua](mailto:mykyta.kyt@nure.ua)

## ABSTRACT

The growing demand for telemedicine services necessitates innovative solutions for monitoring patients with movement disorders. The proposed mobile video surveillance system integrates computer vision, deep learning, and intelligent analysis methods to diagnose human movement anomalies accurately in real-time. The system's key components are algorithms for detecting and tracking human movements, contextual analysis, and a decision-making system. The system makes decisions about the detection or possibility of incidents, such as falls, and sends notifications to users via a mobile application. The integration of IoT sensors enables the processing of physiological data (heart rate, activity), which increases the accuracy of diagnostics and expands the system's range of applications. The proposed modular architecture of the system provides scalability and adaptability to different operating conditions. The system meets privacy standards through the use of encryption and multi-level authentication. Our proposed system ensures effective monitoring of patients with motor disorders, detects incidents promptly, and sends alerts, improving the quality of telemedicine services.

## KEYWORDS

mobile health, patient monitoring, diagnostics of human movement abnormalities, computer vision, video stream processing, contextual analysis, intelligent analysis

## 1 INTRODUCTION

The modern healthcare system faces numerous challenges, such as an aging population, an increase in chronic diseases, including those associated with impaired motor functions, and a growing demand for remote monitoring and telemedicine services [1]. Innovative health monitoring technologies based on the analysis of patient's movement patterns, especially using mobile platforms, are an important area of medical development, as they allow monitoring patients from a distance and automatically diagnosing and reporting abnormalities in movement [2].

Scientific approaches to physical condition monitoring using artificial intelligence (AI) and computer vision technologies demonstrate significant potential

Kyt, M., Yesilevskyi, V. (2025). Development of a Mobile Video Platform for Diagnosing Human Motor Anomalies. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(11), pp. 28–44. <https://doi.org/10.3991/ijoe.v21i11.56897>

Article submitted 2025-06-05. Revision uploaded 2025-07-11. Final acceptance 2025-07-11.

© 2025 by the authors of this article. Published under CC-BY.

for automated anomaly detection. For example, the wearable devices and sensors reviewed by Patel et al. [3] are actively used in rehabilitation programs to assess patients' physical parameters. However, such devices are limited in functionality and cannot fully provide the reliable video monitoring required to diagnose complex movement anomalies. Similar limitations are also inherent in other systems: for example, MediaPipe [4], widely used to track posture and movements, does not allow detection of subtle deviations in exercise performance, which are critical for accurate diagnosis and rehabilitation.

Our proposed mobile video platform system for detecting movement anomalies (see Figure 1) differs from existing solutions' ability to integrate multiple technologies. It combines segmentation and precise motion monitoring capabilities based on deep learning, computer vision technologies, and mobile platforms, allowing for real-time data acquisition and continuous monitoring of the patient's physical condition. Unlike existing platforms that focus on individual health indicators, this system provides comprehensive monitoring, which allows for the detection of hidden patterns and movement anomalies that are critical for the complete rehabilitation of patients [5].



**Fig. 1.** Illustration of the use of a video platform for the diagnosis of motor anomalies (Figure generated with the aid of AI)

Thus, the main goal of this work is to propose a new approach to creating an integrated mobile platform for tracking and diagnosing movement anomalies, which includes a modular structure that can connect additional sensors and IoT devices. This solution creates new opportunities for improving the healthcare system due to its high accuracy, adaptability, and expandability.

## 2 RELATED STUDIES

This section reviews recent advancements that directly inform the development of a mobile video platform for diagnosing movement anomalies. The selection of technologies and methods is grounded in current trends in digital health, remote monitoring, and the rapid evolution of computer vision and deep learning in medical diagnostics.

Modern telemedicine and rehabilitation increasingly rely on real-time video analysis to detect abnormal movement patterns. Vision transformers (ViT) [6], a state-of-the-art deep learning architecture, have demonstrated superior performance in

analyzing complex video streams. These models enable the automatic recognition of motion sequences and real-time anomaly detection, which are essential for precise and automated movement diagnosis.

Another cornerstone technology is MediaPipe Pose Estimation [7], which provides robust tracking of key body landmarks—such as joints and limbs—using only a mobile device camera. Accurate pose estimation is fundamental for identifying deviations from normal movement, enabling clinicians and automated systems to assess patient mobility and detect even subtle anomalies critical for early intervention and rehabilitation planning.

The Segment Anything Model (SAM) [8] further enhances diagnostic accuracy by automatically segmenting and highlighting body regions within video frames. This technology supports fine-grained analysis of movement, ensuring that exercise form and rehabilitation protocols are precisely followed and allowing the system to quantify specific body part engagement or deviation.

Integration with IoT devices expands the diagnostic capabilities of the platform. Wearable and ambient sensors provide complementary physiological data—such as heart rate, body orientation, and activity levels—that, when fused with video analytics, create a comprehensive profile of the user's physical state [9]. This multimodal approach increases the sensitivity and specificity of movement anomaly detection, capturing complex patterns that may indicate underlying health issues.

The proposed mobile video platform is designed with an adaptive, modular architecture, supporting seamless integration of new algorithms, sensors, and hardware components as technology evolves [10]. This flexibility ensures long-term scalability and enables continuous updates in response to emerging clinical and technical requirements.

Collectively, these technologies enable the creation of a mobile video platform that delivers precise, real-time diagnosis of movement anomalies. By leveraging advances in computer vision, deep learning, and IoT integration, the system offers effective, scalable, and user-friendly solutions for telemedicine, rehabilitation, and ongoing health monitoring.

### 3 METHODOLOGY

#### 3.1 Architecture of the data collection and processing system

The developed mobile video platform for detecting movement anomalies is based on modern computer vision, deep learning, and big data processing technologies. It involves a multi-level process of data collection, processing, and analysis, which allows the detection of anomalies in patient behavior in real-time, taking into account the context and specific environmental conditions. The main steps of the process of collecting, processing, and outputting information, as well as the key components of the architecture that implement the functionality of the proposed platform, are described below.

##### **Step 1: Integrate indoor cameras and the mobile app**

1. Stationary cameras: Installing high-resolution cameras in controlled areas (e.g., hospitals or nursing homes) ensures constant monitoring and the ability to continuously transmit video streams to a server for further analysis [11]. Each camera has a stable network connection, providing reliable data transmission to the processing server.

2. Mobile application: The mobile app acts as an interface for healthcare professionals and relatives, providing data visualization, real-time alerts, and access to event history. The application is integrated with the server via the REST API, which ensures instant data synchronization and sends push notifications in case of critical events, such as falls or abnormal movements [12].
3. Central data processing server: Based on FastAPI or Flask, the server acts as a central node that receives video streams, coordinates data processing, and interacts with a decision-making model (Llama 3.2) to analyze falls. The server gets real-time video frames, processes them, and manages data storage for further analysis [13].

### **Step 2: Detecting people and estimating body position**

1. Video stream processing and frame extraction: The system is optimized to allocate video frames at regular intervals (for example, every 0.5 seconds), which reduces the computational load while maintaining high analysis speed. This is achieved by using a frame buffering algorithm on the server.
2. Detecting people with YOLO: To identify faces in a video stream, the YOLO model [14] (e.g., YOLOv8) is used to find people in each frame and select bounding boxes for further processing. YOLO works effectively with dynamic scenes and minimizes false positives in cases with many objects.
3. MediaPipe for assessing body posture and position: For each person detected by YOLO, MediaPipe is used to identify key body points such as shoulders, knees, and hips. MediaPipe can identify abnormal postures and rapid movements that could indicate falls or other potentially dangerous events. Using these key points, the system flags as “fall candidate” footage where unusual postures or rapid changes in movement are detected.

### **Step 3: Contextual analysis with segmentation and action recognition**

1. Segment anything model: For footage marked as a “fall candidate,” the SAM model is used to accurately segment the detected person in the context of the overall scene. This allows for a more detailed outline of the body shape, even when the person is close to other objects, such as furniture or medical equipment, reducing false positives.
2. Contextual analysis with CLIP: To improve scene interpretation, the CLIP [15] model is used, which analyzes frames with “fall candidates” to determine the context of the situation. For example, CLIP can help recognize whether a person is near a staircase, bed, or other object that potentially increases the risk of falling. This allows for objects in the environment to be considered and contextual data to be added to the analysis to minimize false positives.

### **Step 4: Making decisions about falling with Llama 3.2**

1. Data aggregation for Llama 3.2: The system collects key parameters for analysis using the Llama 3.2 model [16]: bounding box coordinates, key body position points, segmentation masks, and contextual data obtained from CLIP. This data is structured into a query sent to Llama 3.2 to decide the probability of falling. For example, the query may look like this: “A person was found lying near the stairs after a sudden change in posture. Is this likely to be a fall?” Llama 3.2

processes the data and returns a probabilistic estimate or textual description of whether a fall has occurred [9].

2. Activation of notifications based on Llama analysis 3.2: If Llama 3.2 estimates a high fall probability, the server generates a notification containing a brief description of the event, for example, "Fall detected near stairs." This notification and a snapshot of the footage are transmitted to the mobile application.

**Step 5: User interface and notification management via the mobile app**

1. Display notifications: The mobile app provides an interface for viewing alerts, which displays the frame, location of the incident, and probability of fall. Users (e.g., caregivers or medical professionals) can view the alert, acknowledge it, or cancel it, allowing feedback to further improve the model’s accuracy.
2. Push notifications and integration with telemedicine: Notifications are sent to guardians or emergency contacts with a brief event description. The data can be automatically transferred to telemedicine systems for immediate intervention or remote consultation if the event is confirmed.

**Step 6: Logging events and continuous improvement**

1. Event logging and feedback processing: All events are stored in logs, including frames, Llama 3.2 responses, and user feedback. This allows you to analyze model accuracy and use the collected data to train or retrain models to improve accuracy.
2. Improving the model: The system can adapt thresholds for detecting falls and other parameters based on user feedback. Additional conditions or refinements to the scene analysis can improve accuracy and adjust the model to new operating conditions [13].

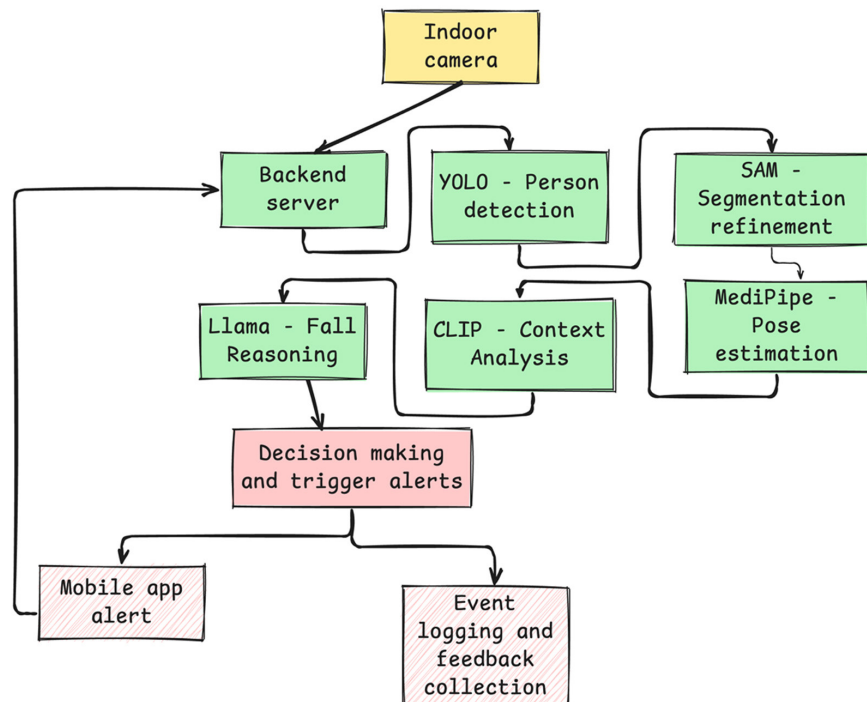


Fig. 2. Pipeline video platform for the diagnosis of motor anomalies

The video platform for the diagnosis of motor abnormalities, shown in Figure 2, includes the following technologies:

- Video processing and object detection: YOLOv8 for face detection, MediaPipe for body pose estimation.
- Segmentation and contextual analysis: SAM for object segmentation, CLIP for contextual classification (optional).
- Decision-making: Llama 3.2 to analyze data and conclude the falling probability.
- Server processing: FastAPI or Flask for frame processing, notification management, and mobile app interaction.
- Mobile application: A platform for receiving notifications, user interaction, and feedback management.

This architecture provides a flexible and accurate process for monitoring user health, combining computer vision, deep learning, and contextual analysis technologies to achieve high reliability and adaptability of the system.

### 3.2 Conceptual architecture of a mobile video platform for the diagnosis of motor anomalies

The conceptual architecture of the mobile video platform for diagnosing movement anomalies is based on a combination of several integrated modules, each of which performs specific functions in the system's overall structure. This approach ensures a high level of scalability, the ability to incorporate additional modules, and adaptation to different use cases. Below is a detailed description of the leading architectural components and levels of interaction.

The primary levels of conceptual architecture:

- **Level of data collection**
  - **High-resolution cameras** provide continuous video surveillance in controlled areas such as patient rooms or medical offices. The cameras are connected to a network and transmit data in real-time to a processing server [17].
  - **IoT sensors** complement cameras by recording critical physiological parameters of the user, such as heart rate and activity level, which allows the system to perform comprehensive monitoring [18].
- **Level of data pre-processing**
  - **Face detection module (YOLO):** The YOLO algorithm identifies the presence of faces in the frame and creates bounding boxes around each detected object. This reduces computational overhead because further analysis is focused only on the selected areas of the frame.
  - **Posture assessment with MediaPipe:** After detecting the faces in the frame, MediaPipe identifies the main key points of the body (joints, limbs) and performs a fundamental posture analysis, which allows you to identify anomalies in body position and mark potential falls.
- **Level of contextual analysis**
  - **Segment anything model** performs segmentation of frames where potential falls are marked. This provides more detailed identification and allows you to distinguish body contours from the background or other objects. SAM improves accuracy in environments with multiple objects, reducing the number of false positives.

- **Contextual analysis with CLIP:** CLIP performs a contextual analysis of the scene to determine if a person is near potentially dangerous objects (stairs, furniture, etc.). This helps to consider the environment to detect falls more accurately, increasing the solution's reliability.
- **The level of decision-making based on AI**
  - **The Llama 3.2 model** receives structured data from the previous levels: bounding boxes, key points of body position, segmented masks, and contextual information. Based on this, a query is generated for the model, which contains a description of the detected situation and the probability of a fall. Llama 3.2 analyzes this information, returning a likelihood or a textual conclusion about the possibility of an incident.
  - **Automatic notification:** If Llama 3.2 detects a high probability of a fall, the system generates an alert. This alert is sent to the mobile app, allowing users to view the incident footage and relevant details.
- **Level of user interface and feedback system**
  - **The mobile app** serves as the primary interface for viewing alerts in real-time. Users can view incident footage, confirm or cancel alerts, and provide feedback. This allows data to be collected to improve the system further.
  - **Push notifications and integration with emergency services:** The app supports sending push notifications to emergency contacts and medical personnel when a critical incident is confirmed. This makes it possible to respond quickly to falls or other dangerous situations.
- **Level of continuous improvement**
  - **Event logging and data analysis:** All events, including Llama 3.2 responses and user feedback, are stored in logs, allowing you to analyze system accuracy, reduce false positives, and optimize processing algorithms.
  - **Adaptation of model parameters:** Based on user feedback, the system dynamically adjusts parameters such as thresholds for anomaly detection, which allows it to adapt to new operating conditions and improve system accuracy continuously [19].

### 3.3 Integration with mobile applications and IoT devices

Integrating a mobile video platform with mobile applications and IoT devices is key to creating a comprehensive approach to diagnosing movement anomalies. This integration provides continuous monitoring, data collection from multiple sources, and the ability to respond immediately to critical situations. Below, we describe the integration mechanisms and their role in ensuring the platform's efficiency.

- **Communication with the mobile application:**
  - The mobile application is the primary means of interaction between the system and users, including healthcare professionals or family members of patients. The application is synchronized with the processing server via an API (e.g., REST API), allowing you to receive and transmit data quickly.
  - App users can receive real-time notifications of detected falls or movement anomalies, view footage of incidents, and confirm or cancel alerts. This feedback allows the system to improve its accuracy based on real-world feedback.
- **Integration of IoT sensors:**
  - IoT sensors such as heart rate monitors, pedometers, and accelerometers are connected to the system via specialized data transfer protocols such as

MQTT or HTTP. These sensors transmit the user's physiological parameters to the processing server, combining them with video data to create a complete picture of the patient's physical condition [20].

- Physiological indicators obtained from IoT sensors add a layer of contextual analysis, allowing the system to detect motor anomalies and assess the user's overall health. For example, a combination of low motor activity and an elevated heart rate can signal a critical situation, such as a heart attack or other cardiovascular problems [21].
- **Data transmission infrastructure:**
  - The system uses protocols that support high-speed transmission and low latency to ensure continuous data exchange. MQTT ensures reliable data transmission even in case of poor connection, as the protocol can store messages until the connection is restored. HTTP provides stable and secure data exchange, especially for transmitting physiological indicators [22].
  - Thanks to these technologies, the platform can support continuous real-time monitoring, which is especially important for immediate response to potential incidents.
- **Notification and feedback system:**
  - The notification system is designed to inform users about critical events in real-time. Push notifications are sent to the mobile application with a description of the incident and appropriate recommendations. For example, the notification may contain the text "A fall was detected near the stairs" along with a snapshot of the incident, which allows for a quick assessment of the situation.
  - In addition, users can provide feedback directly through the app, which allows them to confirm or deny the detected incident. This feedback is sent back to the processing server, which is used to improve the algorithms further and reduce the number of false positives.

Integrating mobile applications and IoT devices significantly increases the video platform's efficiency, allowing users to monitor and analyze patients' conditions in real-time. This approach also increases the convenience and accessibility of the system for medical staff and patient's relatives, which positively impacts the quality and timeliness of services provided.

### 3.4 Notification and feedback system

When a motion anomaly is detected, the notification and feedback system automatically generates an alert sent to the user's mobile application. The instant messaging system ensures immediate notification of responsible persons, critical for timely response to potentially dangerous incidents.

Interactive feedback in the mobile app allows users to confirm or cancel alerts, significantly improving the system's accuracy, especially in reducing the number of false positives. This information is used to train recognition algorithms.

In cases where the seriousness of the incident is confirmed (for example, if the fall resulted in an injury), the system automatically initiates an escalation, sending a notification to the specified emergency contacts.

The alert system also provides analytics and reporting capabilities. Regular analysis of feedback allows the platform to improve the accuracy of critical event detection by adapting thresholds or adjusting parameters for specific monitoring conditions.

### 3.5 Data protection and privacy

Data protection and user privacy are key aspects of the mobile video platform used to diagnose movement anomalies. To ensure the safety of personal and medical data, modern encryption technologies and multi-level security mechanisms are used to guarantee the secure exchange of information between different platform components.

The platform implements data encryption both during transmission and storage. Data is transferred between devices via secure communication protocols such as HTTPS and MQTT with TLS support, which protects data from interception during transmission. All video and medical data collected by the platform is encrypted using AES-256 algorithms, a standard for protecting confidential information, and complies with international security standards [23].

Access to the system is restricted through multi-level authentication, including two-factor authentication (2FA) for users with administrative privileges and health-care professionals. This ensures protection against unauthorized access, particularly in device loss or theft. In addition, the system periodically updates access tokens to provide an additional layer of protection against malicious attacks.

Privacy is also maintained through role-based access restriction policies. This segmentation allows you to provide access only to the necessary information, reducing the risk of a privacy breach.

The system also complies with international data protection standards, including the General Data Protection Regulation of the European Union (GDPR) and the Health Insurance Portability and Accountability Act of the United States (HIPAA). This ensures that users' privacy rights are respected, as well as the platform's obligation to report security incidents [19].

Thus, the platform utilizes advanced security and privacy practices to ensure the safety of user data at all levels. All measures, including encryption, multi-level authentication, access restriction policies, and compliance with international standards, make the platform reliable and protected from external threats.

### 3.6 Implementation of the system at the level of subsystem interaction

The implementation of the mobile video platform system for the diagnosis of motor anomalies includes the integration of subsystems within the framework of a pilot project that demonstrates the ability of various components to interact and coordinate to achieve high diagnostic accuracy. The pilot implementation is designed to test the effectiveness of the main modules at the integration level by simulating the interaction between them.

At the video data processing level, integration with pose detection and estimation algorithms such as YOLO and MediaPipe is used. The test video frames are transferred to the server, where these algorithms detect faces and estimate their positions.

For a more detailed analysis of the subsystem's interaction, a segmentation module was implemented using the SAM model, which provides outline detection of objects in the frame. Based on the data obtained, the contextual analysis module, built using the CLIP model, identifies objects around the patient, including potentially dangerous objects such as stairs or a bed. This helps to improve accuracy and reduce the risk of false positives by considering the context of each situation.

As part of the pilot project, we also implemented interaction with the decision-making system based on Llama 3.2, enabling probabilistic event estimates to form. The system aggregates data from the previous modules and generates structured queries to Llama 3.2, which interprets the data and returns an estimate of the probability of a critical event.

The pilot project's alert system integrates with a mobile application where users can receive notifications of potential incidents in real-time.

Thus, the implementation of the pilot project demonstrates the effectiveness of the interaction of subsystems and confirms that the proposed system can support the actual operation of a mobile video platform for the diagnosis of motor anomalies. This creates an opportunity for further scaling the system, including integrating additional functions and modules into a full version of the platform for clinical use.

### 3.7 Data collection, annotation, and technical specifications

All data used in this study were collected by the authors using stationary high-resolution cameras (Hikvision DS-2CD2043G0-I) installed at a height of 2.2 meters in controlled clinical environments, including hospital wards and rehabilitation centers. Each camera was configured to record continuous Full HD (1920 × 1080 pixels) video streams at 25 frames per second, ensuring an unobstructed and comprehensive view of the monitored area. The data collection protocol encompassed both routine daily movements and staged abnormal events, resulting in over 400 hours of video footage acquired specifically for this study.

Manual annotation of the collected video data was performed using the CVAT tool. Each frame containing a person was labeled by marking bounding boxes in the YOLO format. To maintain high annotation quality, a two-stage validation protocol was implemented: all annotations were independently reviewed by a second annotator, and frames with intersection over union (IoU) values below 0.8 were re-annotated. Periodic inter-annotator agreement checks were also performed to ensure annotation consistency and reliability.

The technical setup for data collection included the use of Hikvision DS-2CD2043G0-I cameras, a local network with continuous power supply, and dedicated video storage infrastructure. Annotation tasks were carried out on a workstation equipped with 32 GB RAM and an Intel Core i7 processor. This self-collected and rigorously annotated dataset forms a reliable basis for all subsequent analysis and evaluation conducted in this study.

## 4 RESULTS

The proposed mobile video platform system for diagnosing motor anomalies demonstrates significant potential through the integration of advanced computer vision techniques, deep learning models, and automated contextual analysis. The system's primary goal is to provide accurate, continuous monitoring of patients' physical conditions, enabling the detection of anomalous movement patterns, timely recognition of potentially dangerous events such as falls, and rapid communication of alerts to medical personnel or caregivers. The architecture of the system is modular and scalable, ensuring flexibility for future enhancements and adaptations to diverse clinical and home environments.

The platform is composed of multiple sequential technological stages. At the first level of video data processing, the system employs the YOLO object detection framework to identify human subjects within video frames and to generate precise bounding boxes around detected individuals. This step is fundamental for isolating regions of interest and significantly improves the computational efficiency and accuracy of all subsequent processing stages by focusing analytical resources exclusively on relevant areas of each frame.

Once individuals are detected, the system performs fine-grained pose analysis using the MediaPipe model, which identifies and tracks anatomical keypoints on the human body, including the shoulders, elbows, hips, knees, and ankles. MediaPipe operates through a lightweight convolutional neural network architecture capable of delivering high-precision 3D landmark estimation and per-joint confidence scores in real-time. This capability allows the system to capture subtle posture transitions and rapid shifts in body orientation that may signify abnormal events such as falls.

Further analysis is performed through the SAM, which produces pixel-level segmentation masks for each detected person. SAM excels at differentiating human silhouettes from background elements, even in visually cluttered scenes containing overlapping objects such as furniture or other individuals. This level of detail allows the system to quantify the spatial geometry of the human body accurately and to maintain consistent person tracking across frames.

Contextual analysis is a critical aspect of the platform, provided by the CLIP model. CLIP processes visual information from the broader scene surrounding each detected person, generating semantic embeddings that enable the recognition of environmental objects and spatial context. For example, CLIP can identify objects such as chairs, beds, tables, or structural hazards such as stairs. This contextual awareness is vital because falls that occur near hazardous objects pose increased risk and therefore demand higher prioritization in alerting logic. When CLIP identifies a potentially dangerous environment, the system can adjust thresholds for fall detection sensitivity, reducing the likelihood of overlooking critical incidents.

The decision-making layer of the platform is driven by the Llama 3.2 large language model, which serves as the system's integrative reasoning component. Unlike systems reliant purely on rule-based logic, the Llama 3.2 model processes structured, multi-modal data inputs collected from earlier stages—including pose transitions, movement velocities, durations of postural changes, spatial segmentation outputs, and environmental object contexts—to generate probabilistic assessments of whether a detected event represents a fall or normal activity. The system utilizes a fixed prompt template for communication with the language model. Into this prompt, the pipeline automatically inserts numerical and semantic parameters derived from the earlier detection and analysis stages. This approach ensures uniformity in how data is presented for AI reasoning, eliminates manual query crafting, and enables reproducible, deterministic decision-making. The output from Llama 3.2 consists of a categorical response—either “Fall detected” or “Normal activity”—thereby simplifying the downstream process of triggering notifications.

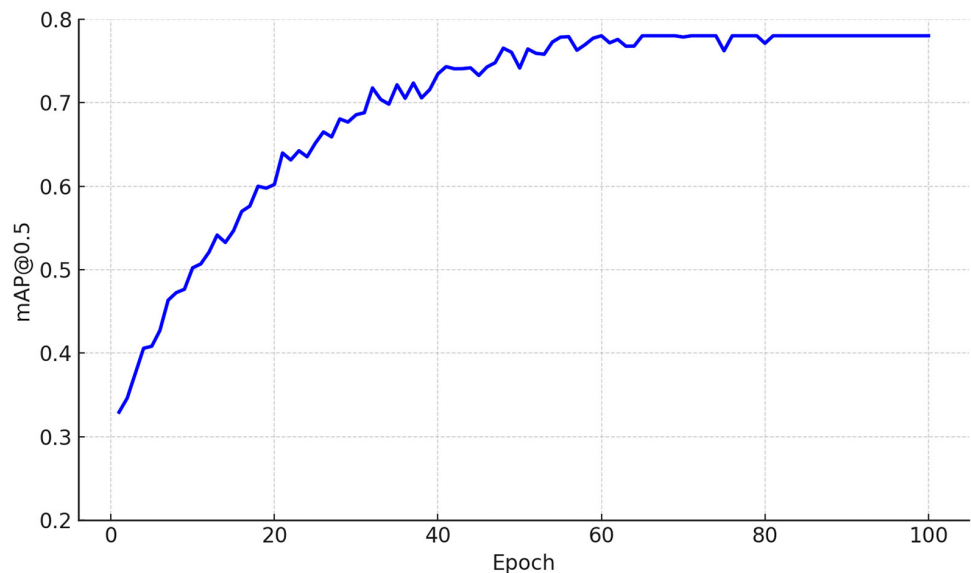
The notification and feedback system represents a crucial operational layer of the platform, delivering real-time alerts to connected mobile devices. In this evaluation, the system was tested exclusively on iOS devices to verify its integration and to assess the practical latency of end-to-end processing. When a fall is detected, the mobile application receives a notification that includes relevant contextual information and a visual snapshot of the incident. Users, such as healthcare professionals or family members, can then confirm the alert or dismiss it as a false positive. This

interactive feedback mechanism not only improves the trustworthiness of the platform but also contributes valuable data for model retraining, enabling continuous improvement in detection accuracy over time.

To quantify system performance, a preliminary experimental evaluation was conducted. The experiments were performed in a controlled indoor environment with carefully managed illumination to minimize variations in visual input quality. Two high-resolution Hikvision DS-2CD2043G0-I IP cameras were installed in opposite corners of the test room to provide overlapping perspectives and mitigate occlusion effects. A total of five healthy volunteers, three male and two female, aged between 25 and 35 years, participated in the data collection sessions. Each subject was instructed to perform two categories of activities: (1) simulated falls, characterized by abrupt posture transitions from vertical to horizontal states, and (2) normal daily activities such as walking, turning, sitting down gently, and standing still. These diverse motion sequences ensured the system would encounter both hazardous events and benign movement patterns.

The video data collected during the sessions resulted in 2000 individual frames, each manually annotated using the LabelImg tool. Bounding boxes around individuals were created and encoded in YOLO format, specifying class labels and normalized coordinate values. This annotated dataset was randomly partitioned into training (80%), validation (10%), and test (10%) subsets to enable objective performance measurement and to mitigate risks of model overfitting.

The YOLOv8 detection model was fine-tuned on the collected data. The training process was conducted over 100 epochs, utilizing a learning rate of 0.001, momentum of 0.937, weight decay of  $5e-4$ , and a batch size of 16 on an NVIDIA RTX 3060 GPU. Despite the modest dataset size, the YOLOv8 model achieved a mean Average Precision (mAP@0.5) of approximately 0.78 on the test set, indicating encouraging yet preliminary detection capabilities. Figure 3 depicts the evolution of mAP across training epochs, confirming convergence stability.



**Fig. 3.** YOLOv8 detection accuracy (mAP@0.5) across epochs

Following integration of all system modules, the platform was evaluated on an independent test subset containing 15 simulated fall events and 20 normal activity cases. The preliminary detection results are summarized in Table 1.

**Table 1.** Preliminary detection results of the mobile video platform

Scenario	Total Cases	Correct Detections	False Positives	Detection Rate (%)
Simulated Falls	15	11	4	73%
Normal Activities	20	16	4	80%
<b>Overall Accuracy</b>	35	27	8	77%

These preliminary results demonstrate that the system is capable of integrating advanced object detection, pose estimation, semantic segmentation, contextual scene analysis, and AI-driven decision-making into a unified framework. However, the current overall accuracy of approximately 77% reflects the system's developmental stage. Several areas have been identified for future improvement, including the expansion of the training dataset to encompass more diverse individuals and scenarios, further fine-tuning of detection and segmentation models, and enhanced domain adaptation techniques to improve robustness across different environments and lighting conditions.

Despite these current limitations, the experimental validation underscores the feasibility and promise of the proposed mobile video platform. The modular architecture and integration of state-of-the-art computer vision and AI technologies lay a strong foundation for achieving highly reliable and adaptive monitoring solutions. Continued development and iterative testing will be critical to achieving clinical-grade performance suitable for deployment in telemedicine, rehabilitation monitoring, and remote patient care applications.

## 5 DISCUSSION

### 5.1 Importance for the healthcare system

The implementation of an automated mobile video platform for movement anomaly detection offers tangible benefits to both healthcare providers and patients. By enabling uninterrupted, real-time observation of individuals, this technology addresses a critical gap in the current continuum of care—particularly for populations at heightened risk of mobility-related incidents, such as the elderly or those with neurological and musculoskeletal disorders.

Integrating this platform into care facilities and patients' homes introduces a paradigm shift toward proactive, rather than purely reactive, healthcare. The ability to remotely monitor patients without the need for constant in-person supervision increases safety while preserving patient autonomy and dignity. Early identification of atypical movement patterns or falls enables faster interventions, reducing the likelihood of complications, lengthy hospitalizations, or long-term disability.

From a systemic perspective, such digital solutions contribute to the evolution of patient-centered healthcare models by supporting aging-in-place initiatives and individualized care plans. Enhanced monitoring facilitates more precise risk stratification and personalized rehabilitation programs, optimizing resource allocation and improving clinical outcomes.

In addition, widespread adoption of video-based monitoring can lead to cost savings for healthcare systems. Automated detection and timely alerts decrease the reliance on labor-intensive manual observation and enable more efficient allocation of medical staff, ultimately lowering operational expenses.

Finally, the incorporation of advanced analytics and IoT sensor data within the platform paves the way for data-driven, adaptive healthcare strategies. Over time, longitudinal movement data collected through this system can support large-scale research, inform public health strategies, and accelerate innovation in telemedicine and remote rehabilitation.

## 5.2 Challenges and limitations

The practical deployment of the proposed mobile video platform in real-world healthcare environments entails several notable challenges and limitations that must be acknowledged and systematically addressed.

One of the primary obstacles lies in the system's reliance on robust technological infrastructure. Accurate anomaly detection and timely alert generation depend on continuous high-quality video input and stable network connectivity. Inconsistent internet access, variable bandwidth, or temporary loss of connection can lead to gaps in monitoring coverage, delayed data transmission, or even missed critical events. The quality of detection is further influenced by camera resolution, placement, and the physical characteristics of the monitored environment. Suboptimal angles, obstructed views, or insufficient lighting can significantly degrade recognition accuracy and increase the likelihood of both false positives and false negatives.

Another challenge concerns the generalizability and reliability of automated detection algorithms outside controlled experimental settings. Although the system demonstrated promising performance in preliminary trials, its sensitivity and specificity may be affected by factors such as unusual patient movements, environmental clutter, or rapid changes in the visual scene. Scenarios involving multiple individuals, pets, or the presence of complex backgrounds can further complicate accurate event classification. To mitigate these issues, ongoing collection and annotation of diverse datasets are necessary, enabling regular retraining and validation of models on new and challenging scenarios.

Maintaining stringent data privacy and security standards is essential, given the sensitive nature of health-related video and sensor data. While encryption, secure data transfer protocols, and multi-factor authentication are implemented, the risk of unauthorized access or data leakage cannot be entirely eliminated, particularly in the face of evolving cyber threats. Regular system audits, proactive vulnerability assessments, and adherence to best practices for data management are imperative to safeguard patient confidentiality and maintain regulatory compliance.

A further technical consideration is the computational load associated with real-time video analysis. Processing high-resolution video streams and integrating data from multiple sources demand significant computational resources, which may not always be available in resource-limited healthcare settings. Balancing detection speed and system scalability with hardware constraints remains a critical engineering challenge, especially when scaling to monitor multiple patients or facilities simultaneously.

Moreover, the platform's performance is influenced by the quality and diversity of the training data used for model development. Inadequate representation of rare events, specific population groups, or unique environmental conditions may limit the system's effectiveness in certain contexts. Ongoing efforts to expand and diversify training datasets, combined with domain adaptation and continual learning approaches, are vital for maintaining high accuracy and minimizing error rates.

In summary, while the mobile video platform demonstrates clear advantages for remote patient monitoring and early detection of movement anomalies, its transition from pilot studies to widespread adoption will require addressing these multifaceted limitations. Continuous technical refinement, expansion of real-world testing, and a commitment to robust data security protocols are essential to ensure that the system achieves reliable, safe, and effective operation in diverse healthcare environments.

## 6 CONCLUSIONS

The proposed mobile video platform for diagnosing movement anomalies represents an innovative approach to monitoring patients' physical condition, providing automated detection of dangerous movement anomalies. The combination of computer vision algorithms, deep learning, and contextual analysis allows the platform to effectively recognize critical situations and inform medical professionals or decision-makers in real-time. The system's main advantages include high adaptability, the ability to integrate into patients' home environments, and the use of AI technologies to make informed decisions. This makes the platform a promising tool for supporting patients who need constant monitoring, including the elderly or people with disabilities.

Future development of the platform may include integrating new sensors to improve the quality of data collection, optimizing algorithms for operation in different lighting conditions, and adapting to the specific needs of individual user groups. Another critical area is improving the mechanisms for protecting confidential information, allowing the platform to meet the highest security standards. The platform also has the potential to expand opportunities in telemedicine and remote care, providing more comprehensive patient support and the ability to monitor their condition effectively.

Thus, the mobile video platform for diagnosing movement abnormalities is an innovative solution for remote patient monitoring that combines high accuracy, adaptability, and ease of use. Its implementation can be an essential step in developing modern patient-centred healthcare, particularly in the context of telemedicine and supporting the quality of life of older people and people with special needs.

## 7 REFERENCES

- [1] X. Zhang, Y. Wang, and J. Li, "Applications of mobile health technologies in posture monitoring: A systematic review," *Sensors*, vol. 21, no. 3, p. 746, 2021. <https://ieeexplore.ieee.org/abstract/document/10772039/>
- [2] T. Shaik *et al.*, "Remote patient monitoring using artificial intelligence: Current state, applications, and challenges," *WIREs Data Mining and Knowledge Discovery*, vol. 13, no. 2, p. e1485, 2023. <https://doi.org/10.1002/widm.1485>
- [3] S. Matuska, M. Paralic, and R. Hudec, "A smart system for sitting posture detection based on force sensors and mobile application," *arXiv preprint arXiv:2209.10806*, 2022. <https://arxiv.org/abs/2209.10806>
- [4] World Health Organization (WHO), "Guidelines on physical activity and sedentary behavior," 2020. <https://www.who.int/publications/i/item/9789240015128>
- [5] Ministry of Health of Ukraine, "Annual report on the state of public health in Ukraine," 2021. <https://moz.gov.ua>

- [6] L. Horn Iwaya, A. Ahmad, and M. A. Babar, "Security and privacy for mHealth and uHealth systems: A systematic mapping study," *arXiv preprint arXiv:2006.12069*, 2020. <https://arxiv.org/abs/2006.12069>
- [7] International Telecommunication Union (ITU), "Measuring digital development: Facts and figures 2021," 2021. <https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx>
- [8] R. Gadhvi, P. Desai, and Siddharth, "PosePilot: An edge-AI solution for posture correction in physical exercises," *arXiv preprint arXiv:2505.19186*, 2025. <https://arxiv.org/abs/2505.19186>
- [9] Global System for Mobile Communications Association (GSMA), "Mobile economy report 2020," 2020. <https://www.gsma.com/mobileeconomy>
- [10] NHS England & NHSX, "Our strategy to digitise, connect and transform health and care," 2021. <https://transform.england.nhs.uk/digitise-connect-transform/our-strategy-to-digitise-connect-and-transform/>
- [11] S. Denning, P. Frank, J. Gray, and L. Eli, "A review of the quality and impact of mobile health apps," *Annual Review of Public Health*, vol. 43, pp. 117–134, 2022. <https://www.annualreviews.org/doi/10.1146/annurev-publhealth-052020-103738>
- [12] M. N. Islam, M. K. Karim, T. T. Inan, and A. K. M. N. Islam, "Investigating usability of mobile health applications in Bangladesh," *BMC Med. Inform. Decis. Mak.*, vol. 20, 2020. <https://doi.org/10.1186/s12911-020-1033-3>
- [13] European Commission, "eHealth – digital services in health and care," 2020. [https://health.ec.europa.eu/ehealth-digital-health-and-care/digital-health-and-care\\_en](https://health.ec.europa.eu/ehealth-digital-health-and-care/digital-health-and-care_en)
- [14] Z. Yang, D. Song, J. Ning, and Z. Wu, "A systematic review: Advancing ergonomic posture risk assessment through the integration of computer vision and machine learning techniques," *IEEE Access*, vol. 12, pp. 180481–180519, 2024. <https://doi.org/10.1109/ACCESS.2024.3509447>
- [15] E. Max-Onakpoya *et al.*, "Augmenting cloud connectivity with opportunistic networks for rural remote patient monitoring," *arXiv preprint arXiv:1905.05342*, 2019. <https://arxiv.org/abs/1905.05342>
- [16] M. Roy, S. J. Minar, P. Dhar, and A. T. M. O. Faruq, "Machine learning applications in health-care: The state of knowledge and future directions," *arXiv preprint arXiv:2307.14067*, 2023. <https://arxiv.org/abs/2307.14067>
- [17] S. Fan, R. C. Jain, and M. S. Kankanhalli, "A comprehensive picture of factors affecting user willingness to use mobile health applications," *arXiv preprint arXiv:2305.05962*, 2023. <https://arxiv.org/abs/2305.05962>
- [18] T. Cory, W. Rieder, and T.-M. Huynh, "A qualitative analysis framework for mHealth privacy practices," *arXiv preprint arXiv:2405.17971*, 2024. <https://arxiv.org/abs/2405.17971>
- [19] C. Del-Valle-Soto *et al.*, "A comprehensive review of behavior change techniques in wearables and IoT: Implications for health and well-being," *Sensors*, vol. 24, no. 8, p. 2429, 2024. <https://doi.org/10.3390/s24082429>
- [20] S. Debnath, "Integrating information technology in healthcare: Recent developments, challenges, and future-prospects for urban and regional health," *arXiv preprint arXiv:2307.16296*, 2023. <https://arxiv.org/abs/2307.16296>
- [21] J. E. Bardram, "The CARP mobile sensing framework: A cross-platform reactive programming framework for digital phenotyping and mHealth," *arXiv preprint arXiv:2006.11904*, 2020. <https://arxiv.org/abs/2006.11904>
- [22] H. D. Wang and M. Ma, "PhysiQ: Off-site quality assessment of exercise in physical therapy," *arXiv preprint arXiv:2211.08245*, 2022. <https://arxiv.org/abs/2211.08245>
- [23] A. Lisowska, S. Wilk, and M. Peleg, "What do you need to consider when designing mobile health intervention?" *arXiv preprint arXiv:2205.10679*, 2022. <https://arxiv.org/abs/2205.10679>

## 8 AUTHORS

**Mykyta Kyt** is an attending PhD student in Applied Mathematics at Kharkiv National University of Radio Electronics (NURE), Kharkiv, Ukraine (E-mail: [mykyta.kyt@nure.ua](mailto:mykyta.kyt@nure.ua)).

**Valentyn Yesilevskyi** is working as an Associate Professor in Department of Applied Mathematics, Kharkiv National University of Radio Electronics (NURE), Kharkiv, Ukraine.