APSIPA Transactions on Signal and Information Processing, 2024, 13, e23

Original Paper

Continual Learning Based Personalized Abnormal Behavior Recognition Alarm System

Hyewon Song[1], Jiwoo Kang[2*] and Taewan Kim[3*]

[1] *The Department of Electrical and Electronic Engineering, Yonsei University*
[2] *Division of Artificial Intelligence Engineering, Sookmyung Women's University*
[3] *The Data Science Major, Dongduk Women's University*

ABSTRACT

The number of people with mental illness is increasing because of stress or environmental influences. They require close monitoring because of their unpredictable behaviors; however, this is challenging given the lack of adequate numbers of medical staff. To overcome this problem, we propose a personalized abnormal behavior recognition alarm system for closed wards. The proposed system utilizes real-time video analysis to detect and track the locations of the patients, enabling recognition of their abnormal behaviors. In addition, new definitions are provided for specific abnormal behaviors that commonly occur in closed wards, with the adaptation of continual learning in the system. This architecture allows the creation of an abnormal behavior dataset while enhancing the recognition accuracy. The average abnormal behavior recognition accuracy with this system is over 92%. According to test results in real hospitals, about 84% of the medical staff were satisfied with the proposed system. Through the proposed alarm system, the staff could implement immediate actions without

*Corresponding author: Jiwoo Kang, jwkang@sookmyung.ac.kr and Taewan Kim, kimtwan21@dongduk.ac.kr.

Received 09 June 2024; revised 05 July 2024; accepted 30 July 2024
ISSN 2048-7703; DOI 10.1561/116.20240033

careful monitoring. By reducing the probability of occurrence of dangerous incidents, the system not only benefits the health of the patients but also enhances the working environment of the medical staff.

Keywords: Abnormal behavior recognition, continual learning, alarm system, DeepStream, mental illness

1 Introduction

Globally, the number of people with mental illnesses is increasing rapidly owing to societal stress, environmental factors, and the use of substances such as drugs [15, 18]. In cases of severe mental illnesses, such as schizophrenia and bipolar disorders, it is necessary to admit patients to closed psychiatric wards for inpatient treatments. The annual number of such patients requiring inpatient care has been increasing steadily; however, there is a shortage of healthcare personnel who provide for them. According to the *Mental Health ATLAS 2020* [14] released by the World Health Organization (WHO) in 2021, the global number of mental healthcare nurses per 100,000 population decreased by 1.3, dropping from 5.1 in 2014 to 3.8 in 2020. Even though closed psychiatric wards have more nursing staff and security personnel than regular wards, many inevitable incidents occur regularly within these protected environments, such as self-harm, harm to others, and falls. It is the best way for medical staff or attendants to check on their patients through CCTVs or directly to reduce such incidents; however, it is impossible to monitor the patient's status every moment. Contradictorily, if these nurses engage in close surveillance, there may be more incidents due to delayed handling of other tasks. Hence, it is necessary for an efficient and closely monitored system to prevent severe incidents with limited medical staff. It would reduce the mental fatigue of the staff and enhance the efficiency of surveillance. Therefore, we propose a real-time abnormal behavior recognition alarm system for early incident detection.

The overall framework of the proposed system is shown in Figure 1. The system involves the detection of patients, prediction of their postures on location, and recognition of their behaviors from the predicted postures to determine indicators of potential danger. Finally, these results are transmitted to medical staff when the patients are in jeopardy. All these processes operate in real-time and are implemented using the NVIDIA DeepStream SDK [5]. It reduces the processing time from an average of 2.2s to an average of 0.95s on a standard desktop, resulting in a reduction of over 1s.

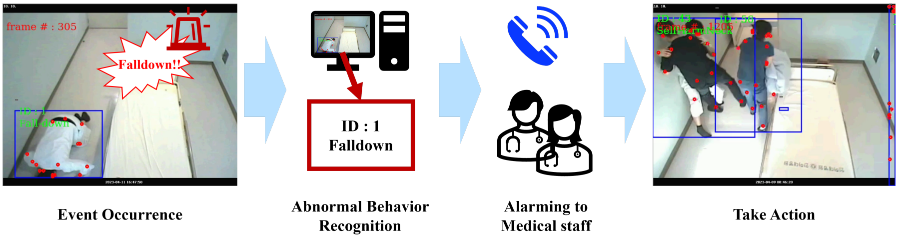


Figure 1: The proposed abnormal behavior recognition alarm system operates as follows: When a patient is in danger, the recognition model identifies the patient and determines the type of abnormal behavior. Subsequently, an alarm is triggered to prompt the medical staff to take necessary actions. The system reduces the workload of the medical staff.

In addition, we defined some of the abnormal behaviors frequently occurring in psychiatric wards [1, 9]. Unlike other common surveillance systems [16, 27, 6] or anomaly detection algorithms [13, 23, 11, 19, 3, 12, 21, 31] typically based on deep-learning architectures, a rule-based behavior recognition is applied to the proposed system. The primary reason for not adopting a common deep-learning method is the absence of a comprehensive abnormal behavior database. Therefore, we established a general definition for each abnormal behavior. The defined abnormal behaviors are thus categorized into four types as follows: SelfHarm, Falldown, Caution, and Hit. To define these abnormal behaviors, the system considers not only joint angles but also postures and temporal continuity.

To address the absence of a comprehensive database on abnormal behaviors, a continual learning method is utilized in the overall system [28, 10]. This approach aimed to overcome the limitations of both the database and rule-based abnormal behavior recognition model. Utilizing the proposed similarity metric, abnormal behavior rules are periodically updated and corresponding data are stored. The average abnormal behavior recognition accuracy of the proposed system is 95.5%. Furthermore, when applied in a real hospital environment for 12 weeks, the system received positive feedback in about 84% of the cases. All of the processes, including database construction and pilot operation, are carried out through collaboration with Inha University Hospital, thereby enhancing the expertise of the proposed system.

The main contributions of our proposed system are summarized as follows:

1. A real-time personalized abnormal behavior recognition alarm system for monitoring mental patients.

2. The definition of four abnormal behaviors (SelfHarm, Falldown, Caution, and Hit) commonly observed in closed wards.

3. A continual learning-based algorithm for creating a unique abnormal behavior database, which improves the performance of the recognition model through updating the behavior rules.

2 Related Works

Unlike the proposed abnormal behavior recognition algorithm, existing abnormality detection algorithms primarily focus on determining whether a behavior is abnormal or not. These algorithms typically rely on datasets with labeled data for general actions, such as the NTU RGB+D dataset [24] or the UTD-MHAD dataset [4], to train models. Subsequently, these models are used to identify abnormal behaviors based on learned patterns.

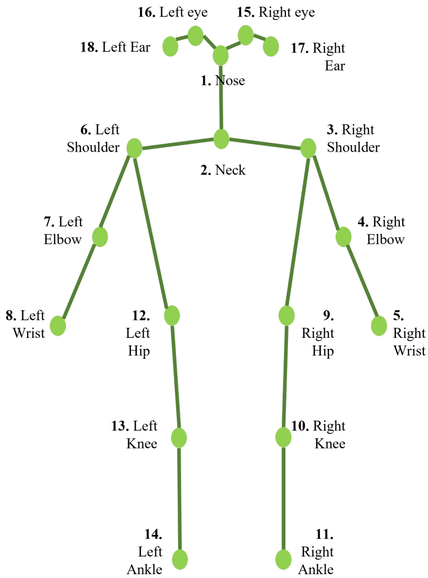
red introduced a novel approach to improve anomaly detection accuracy in video sequences by exploiting the correlation between visual appearance and motion cues. Their method provides a comprehensive understanding of the scene, effectively distinguishing normal activities from anomalies. red, red explored the use of deep neural networks for one-class classification, emphasizing anomaly detection. Their approach leverages the representational power of deep models to encapsulate the normal data distribution effectively, offering a robust technique for detecting rare and unseen events. In red, an unsupervised anomaly detection method using Generative Adversarial Networks (GANs) was proposed. By using GANs to learn the underlying distribution of normal data, their approach effectively identifies deviations indicative of anomalies, providing a promising technique for uncovering hidden anomalies.

In a different approach, red addressed the challenge of anomaly detection in multi-sensor data using an LSTM-based encoder-decoder model. By capturing sequential dependencies among various sensors, their architecture effectively identifies intricate patterns indicative of anomalies, showcasing potential improvements in accuracy for complex scenarios. red also employed Recurrent Neural Networks (RNNs) for multivariate time series anomaly detection. By leveraging temporal relationships in the data, the RNN architecture accommodates the presence of missing values, proving especially promising in capturing contextual information while identifying anomalies, even in scenarios with data gaps.

3 Abnormal Behaviors

3.1 Notations

The definition of a joint as used in this paper is shown in Figure 2(a). It includes 18 joints obtained from the skeleton estimation algorithm. For convenience, the terms are summarized in Figure 2(b). *abc*L(R) means connecting joint *a*, joint *b*, and joint *c* of the left (right) arm or leg. The notation $\theta_{\square}$ represents the angle formed by the two line segments denoted by $\square$. For instance, θ_{SEWL} represents the angle formed by the left shoulder, left elbow, and left wrist.



(a) Definition of Joints

Rep.	Meaning
S	Shoulder
E	Elbow
W	Wrist
H	Hip
N	Knee
A	Ankle
*abc*L	*abc* of Left
*abc*R	*abc* of Right
SEW	Shoulder-Elbow-Wrist
HNA	Hip-Knee-Ankle
SHN	Shoulder-Hip-Knee

(b) Representations

Figure 2: The summary of some notations. (a) the definitions of joints, (b) the terms.

3.2 Types of Abnormal Behaviors

New definitions are introduced for four types of abnormal behaviors: SelfHarm, Falldown, Caution, and Hit. These four abnormal behaviors are the most commonly observed types in closed wards and necessitate alerting the nurses in a real hospital environment. Initial definitions are provided for these abnormal behaviors and implemented in an abnormal behavior recognition algorithm.

3.2.1 SelfHarm

Patients suffering from mental illnesses harm themselves in various ways, such as scratching their arms or legs with their fingernails, hitting their heads with their hands, or making cuts on their wrists or neck with sharp objects like knives or scissors. In this system, SelfHarm was categorized into three specific behaviors: hitting one's head with their hands (Hit Head), choking one's neck with their hands (Strangle Oneself), and causing wounds on one's arms or legs (Injure Arms/Legs). Since a typical closed ward does not permit the patients to bring in objects other than beds and bedding, behaviors involving other objects for SelfHarm were excluded. The examples of three SelfHarm patterns are shown in Figure 3.

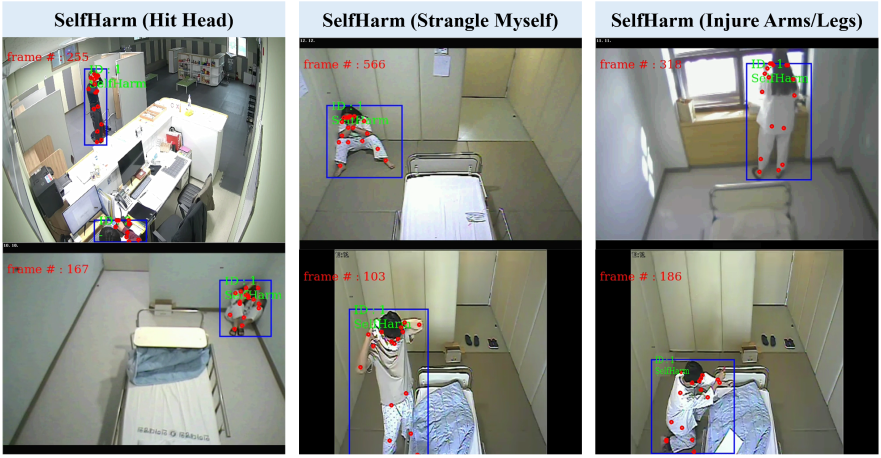


Figure 3: The examples of SelfHarm. There are three types of SelfHarm: Hit Head, Strangle Oneself, and Injure Arms/Legs.

Table 1 summarizes the characteristics of the parameters used for abnormal behavior recognition related to these three types of SelfHarm. Hit Head and Strangle Oneself show similar patterns in terms of posture and, arm angles & positions. These behaviors also involve actions wherein one or both hands are directed toward the neck area. It is different in whether the behavior is maintained for a certain duration or not. Recognizing Injure Arms/Legs is more challenging as there are fewer constraints on the angles and positions of these limbs. Since this behavior is usually performed while sitting in a corner for a long time, it is likely to be classified as SelfHarm or Caution.

3.2.2 Falldown

Falldown is the type of behavior that is considered most common in closed wards. Patients with mental illnesses are often housed in small and solitary rooms with only a bed. Even though patients' behaviors are potentially precarious, in particular, Falldown causes a high risk of injury. It is necessary to monitor closely post-Falldown behaviors. Therefore, there is a need for more

Table 1: Characteristics of SelfHarm(Hit Head, Strangle Oneself, and Injure Arms/Legs). $\square_c$ is the $\square$ of the current frame, and $\square_p$ is the $\square$ of the previous frame. W refers to the y position of the wrist in the image coordinates.

	Hit Head	**Strangle Oneself**	**Injure Arms/Legs**
Posture	Not Lie	Not Lie	Sit
Hands	Near Head	Near Neck	Near Arms/Legs
Angles of Arms	$20 \leq \theta_{SEWL} \leq 90$ or $20 \leq \theta_{SEWR} \leq 90$	$20 \leq \theta_{SEWL} \leq 90$ and $20 \leq \theta_{SEWR} \leq 90$	$\theta_{SEWL} \leq 180$ or $\theta_{SEWR} \leq 180$
Temporal Arms' Angles	$\theta_{SEWLc} \leq \theta_{SEWLp}$ or $\theta_{SEWRc} \leq \theta_{SEWRp}$	$\theta_{SEWLc} \leq \theta_{SEWLp}$ and $\theta_{SEWRc} \leq \theta_{SEWRp}$	-
Positions	$\mathrm{W}_c \leq \mathrm{W}_p$	$\mathrm{W}_c \leq \mathrm{W}_p$	$\mathrm{W}_c \geq \mathrm{W}_p$
Legs	-	-	Almost crouch

precise detection of Falldown compared to other abnormal behaviors. In the proposed system, Falldown was categorized into two types: falling from a chair (Up→Down) and rolling off the bed (Sideways). The examples of Falldown red shown in Figure 4.

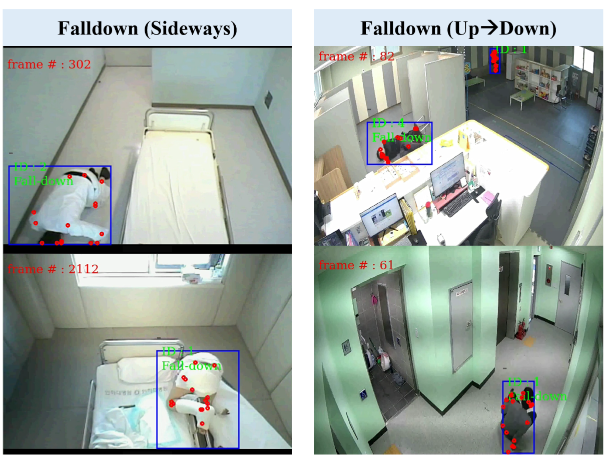


Figure 4: The examples of Falldown. There are two types of Falldown: Up→Down and Sideways.

Table 2 summarizes the two Falldown patterns. Detecting Falldown is quite simple. Falldown was determined by how much the estimated skeleton has moved from top to bottom or side to side in an image. Unlike SelfHarm, the recognition of Falldown is based on tracking the skeleton temporally. In this system, the frame unit was set to 10 frames. This is because checking approximately 10 frames covers a range of speeds as the accuracy of Falldown detection varies significantly with falling speed. Up→Down is recognized when all y-points of the estimated joints increase in the image coordinates. Likewise, Sideways is recognized when all x-points of the estimated joints show significant differences in the image coordinates.

Table 2: Characteristics of Falldown. APy refers to the y points of all skeleton positions in image coordinates. APx refers to the x points of all skeleton positions in the image coordinates. α is set to 0.1*(height of image) and β is set to 0.1*(width of image).

	Up→Down	**Sideways**
Posture	Sit	Sit/Lie
Temporal All Positions	$\mathrm{APy}_p + \alpha \leq \mathrm{APy}_c$	$\mathrm{APx}_c + \beta \leq \mathrm{APx}_p$ or $\mathrm{APx}_p + \beta \leq \mathrm{APx}_c$

3.2.3 Caution

Caution, as defined in the proposed system, shares some patterns with Self-Harm, but it is generally less dangerous and mainly involves staying in one place for a long time. Caution is also derived from the most frequently observed behaviors among patients with mental illnesses in closed wards. The four categories of Caution were defined as follows: hitting the head against the wall (Hit Head), punching the wall (Punch), kicking the wall with the foot (Kick), and staying in a corner for a long period (Stay). Patients admitted to closed wards spend all their time in the room, meaning that they spend significant amounts of time alone. They in states of high distress, anxiety, depression, and similar conditions often exhibit anxiety symptoms, such as picking at their hands or hitting their head against the wall. However, these behaviors have very few movements that are difficult to capture through image processing. Therefore, "Stay" which means that patients stay in one place for a long time was defined as shown in Figure 5.

Table 3 is summarized the types of Caution. As mentioned earlier, the most common type of Caution is staying in one place. Therefore, unlike other abnormal behaviors, separate Intersection Over Union (IOU) limitations were set for the detected boxed between consecutive frames. Hit Head, Punch, and Kick have relatively more movements, so the IOU limitations were set to 0.85,

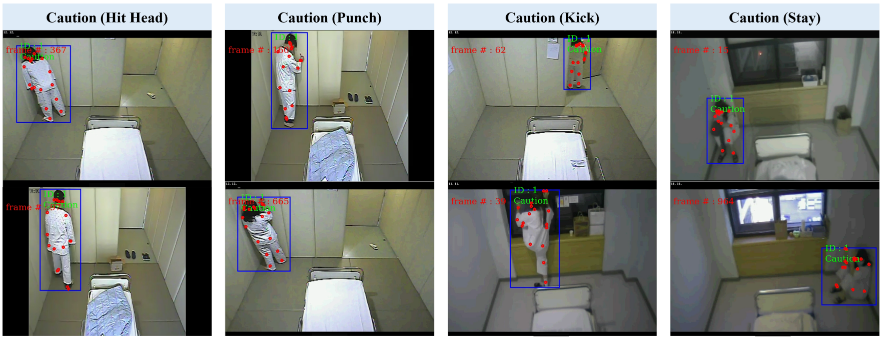


Figure 5: The examples of Caution. There are four types of Caution: Hit Head, Punch, Kick, and Stay.

Table 3: Characteristics of Caution (Hit Head, Punch, Kick, and Stay). All Caution behaviors excluding stay are behaviors against a wall. θ_{HNAL} is the angle formed by connecting the hip, knee, and ankle of the left leg. θ_{HNAR} is the angle formed by connecting the hip, knee, and ankle of the right leg.

	Hit Head	**Punch**
Posture	Stand/Sit	Stand
IOU	0.85	0.8
# of frames	10	10
Angles of Arms	$90 \leq \theta_{SEWL} \leq 180$ or $90 \leq \theta_{SEWR} \leq 180$	$45 \leq \theta_{SEWL} \leq 75$ or $45 \leq \theta_{SEWR} \leq 75$
Temporal Arms' Angles	$\theta_{SEWLc} \leq \theta_{SEWLp}$ or $\theta_{SEWRc} \leq \theta_{SEWRp}$	$\theta_{SEWLc} \leq \theta_{SEWLp}$ or $\theta_{SEWRc} \leq \theta_{SEWRp}$
	Kick	**Stay**
Posture	Stand	Stand/Sit
IOU	0.8	0.9
# of frames	10	20
Angles of Legs	$90 \leq \theta_{HNAL} \leq 180$ or $90 \leq \theta_{HNAR} \leq 180$	-
Temporal Legs' Angles	$\theta_{HNALc} \leq \theta_{HNALp}$ or $\theta_{HNARc} \leq \theta_{HNARp}$	-

0.8, and 0.8, respectively. In contrast, Stay shows minimal movements, so the IOU limitation was set to 0.9. The number of frames considered for the IOU limitations was also set to 20 frames, which is 10 frames more than that used for Hit Head, Punch, and Kick. Although the angles of the arms or legs are also important in Caution, the limitations for these types of Caution were relatively lighter than for SelfHarm because these behaviors have minimal movements.

3.2.4 Hit

The last of the abnormal behaviors defined in the proposed system is Hit. This is the only behavior that occurs when there are two or more people. It refers to the act of striking another person with a fist or foot as shown in Figure 6. Unlike other abnormal behaviors, Hit had a relatively short interval. For this reason, a short frame interval was set to recognize Hit effectively.

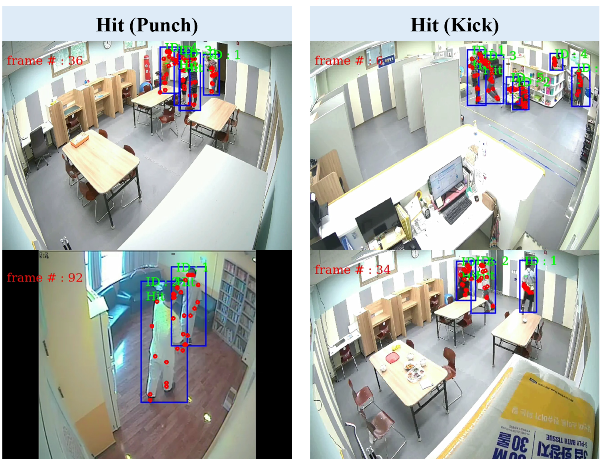


Figure 6: The examples of Hit. There are two types of Hit: Punch and Kick. Compared to other abnormal behaviors, there are two or more people in the same place.

Table 4 summarizes the characteristics of Hit. Each parameter is related to a single person (the aggressor). The definitions of Hit were a combination of the parameters for Caution's Punch and Kick. However, the constraints for each parameter were relatively lower since Hit had more significant movements compared to Caution. In addition, constraints on the distance between the two detected bounding boxes (the aggressor and the victim) were used to determine whether a behavior was Hit. When striking another person with a fist or leg, the corresponding fist or leg is close to or overlaps with the bounding box of the other person. This closeness was represented using γ and δ, where γ was

Table 4: Characteristics of Hit when person 1 is the perpetrator. W_1 and A_1 are the wrist and ankle positions of person 1. B_2 is the x position of person 2's bounding box.

	Hit(Punch)	**Hit(Kick)**
Posture	Stand/Standbend	Stand/Standbend
Distance	$\|W_1 - B_2\| \leq \gamma$	$\|A_1 - B_2\| \leq \delta$
Angles of Arms	$90 \leq \theta_{SEWL} \leq 180$ or $90 \leq \theta_{SEWR} \leq 180$	-
Angles of Legs	-	$90 \leq \theta_{HNAL} \leq 180$ or $90 \leq \theta_{HNAR} \leq 180$
Angles of Body	-	$90 \leq \theta_{SHNL} \leq 150$ or $90 \leq \theta_{SHNR} \leq 150$

set to 0.3 times the length of the first person's arm, and δ was set to 0.3 times the length of their leg. It means that δ and γ were highly dependent on the person 1's body information, not relying on a constant value. For Hit's Kick, the angles of the body were also considered, unlike Caution's Kick. When kicking another person, the range of motion for the leg is much wider than when kicking a wall.

4 Abnormal Behavior Recognition Alarm System

The overall framework of the proposed abnormal behavior recognition alarm system within a closed ward is shown in Figure 7. This system can be broadly categorized into two main components, of which the first is the initial database and defining the initial behavior rules. Before using the system, it is necessary to gather some personal data. In addition, some initial behavior rules are needed, as described in Section 3. The second component entails performing abnormal behavior recognition on video obtained from CCTV. The abnormal behavior recognition process comprises three parts: Human Detection & Skeleton Estimation, Rule-based Abnormal Behavior Recognition, and Continual learning-based behavior rules update. The primary objective of the system is to identify potential incidents accurately and promptly. To achieve rapid and precise incident detection, the whole system is implemented with the DeepStream SDK for efficient video processing. Furthermore, through continual learning, it is able to not only acquire personalized abnormal behavior data but also enhance the overall performance of abnormal behavior recognition.

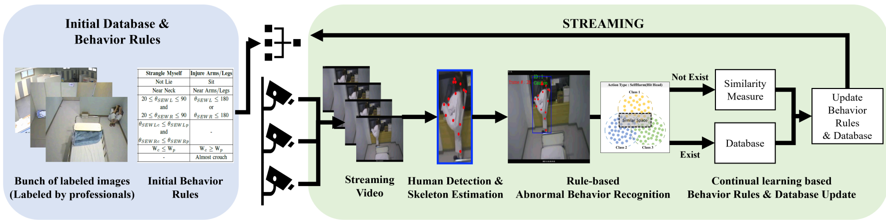


Figure 7: Overall framework of the proposed personalized abnormal behavior recognition algorithm. With the initial database and behavior rules, the proposed system is performed in the order of Human Detection & Skeleton Estimation, Rule-based Abnormal Behavior Recognition, and Continual learning-based behavior rules update.

4.1 DeepStream SDK

DeepStream SDK is a high-performance library provided by NVIDIA and is designed to facilitate easy development of high-throughput video analysis applications that leverage deep learning [5]. It excels in tasks such as object detection, image classification, and instance segmentation-based AI models. With its advanced C++ API and high-performance runtime, it enables rapid integration of GPU-accelerated transcoding and deep-learning inference capabilities, allowing more responsive AI-based services. To ensure swift response times for the risk alerts, the DeepStream SDK was employed. The DeepStream pipeline within the proposed method is shown in Figure 8.

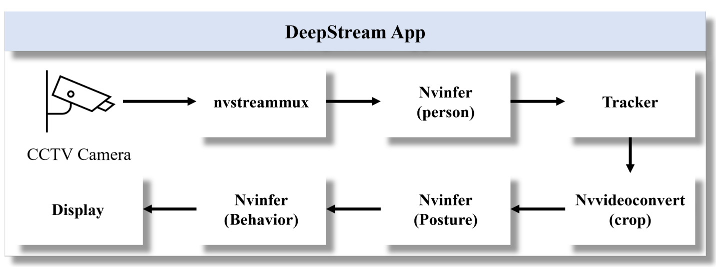


Figure 8: Overall system operation processes in DeepStream.

4.2 Human Detection

The first step in the proposed system is to locate the patient. It captures the positions of patients in each frame from the streaming CCTV videos within the hospital rooms or common areas. We used YOLO v4 [2, 17], a widely used for human detection. In addition to detection, the patients were tracked to allow temporal tracking of their behaviors. Considering the unique characteristics of a closed hospital ward, the proposed system focused more on the accuracy of object detection rather than human detection. The detection threshold was set to 0.4, which is relatively low. In other words, the proposed system paid more attention to whether some objects (including humans) were detected or not rather than the detected object was human. This is because the main goal of the proposed system was to reduce the risk of false negatives to ensure patient safety. Furthermore, in the subsequent processes, such as skeleton estimation, wrongly detected bounding boxes would naturally be filtered out, so the parameters were set to capture as many objects as possible in the video.

4.3 Skeleton Estimation

The skeleton estimation model was based on HRNet [25, 29]. HRNet is a model that achieved state-of-the-art results for human pose estimation in 2019, which is designed for single-person pose estimation. HRNet maintains parallel subnets of various resolutions and continuously exchanges global context and local information through exchange units. Based on these characteristics, it enables more accurate skeleton estimation for bounding boxes of various scales. It not only captured the positions of each joint but also calculated the reliabilities of these joints, which were used as constraints in rule-based abnormal behavior recognition. The results of human detection, tracking, and skeleton estimation within the proposed method are shown in Figure 9.

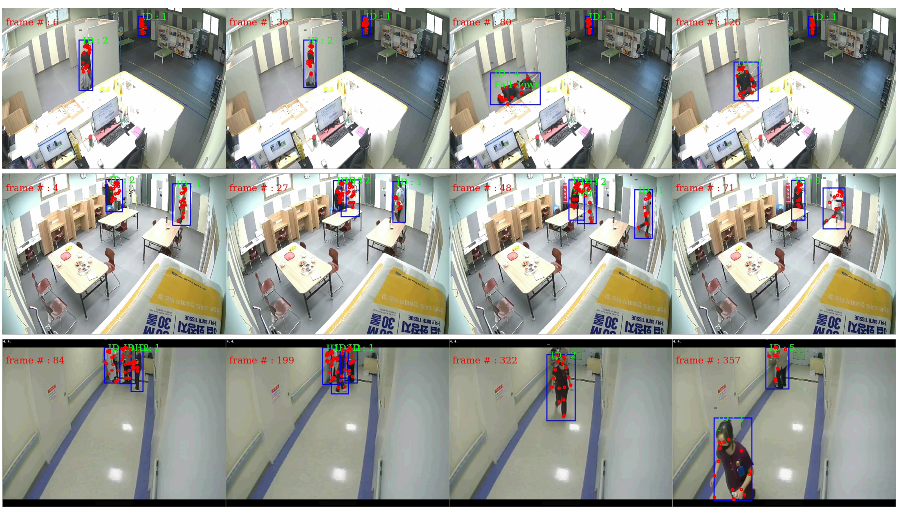


Figure 9: The results of human detection, tracking, and skeleton estimation. The image in each line is excerpted from one video. Wherever there is a human, a bounding box and the estimated skeleton are displayed. In addition, it shows that the assigned ID to each person remains the same even as the frame progresses.

4.4 Rule-Based Behavior Recognition

For abnormal behavior recognition using human skeletons, we adopted the rule-based algorithm [20, 26]. Abnormal behaviors were detected when multiple conditions, such as posture and angles of the arms, match for each specific type of behavior. Before the behavior recognition, all joints should meet the following two conditions:

1. The number of joints with a reliability value of 0.8 or higher should be greater than or equal to 10.
2. When calculating the angles of the joints like θ_{SEW} or θ_{HNA}, the reliabilities of all joints used in the angle calculations should be 0.85 or higher.

Condition 1 underscored the significance of a robust filtering mechanism to ensure the reliability of the estimated skeleton. This mechanism played a pivotal role in eliminating incomplete or untrustworthy skeleton estimations resulting from factors like poor image quality or severe occlusion. As highlighted earlier, the primary objective of the proposed algorithm was to minimize false positives, necessitating the exclusion of unreliable detections based on skeleton estimation reliability. Condition 2 emphasized the reliabilities of the joints involved in the angle calculations and aimed to enhance the accuracy of the angle-based criteria used for abnormal behavior recognition. Joint reliability emerged as a crucial factor, especially in capturing posture details of limbs. Consequently, only joints with a reliability score exceeding 0.85 were considered when calculating angles.

In rule-based behavior recognition system, the execution sequence of code was determined by the significance of the abnormal behavior type. Real-world scenarios often involve the overlap of multiple behavior types, such as a person falling and simultaneously engaging in SelfHarm by choking or hitting their head. In such complex situations, it becomes crucial to detect and address multiple behavior types concurrently. To accommodate the healthcare environment, it is essential to prioritize the detection of behavior types associated with a higher risk of injury. Therefore, the behavior recognition algorithm adhered to a specific order—Falldown, Hit, and SelfHarm. This prioritized sequence ensured that medical staff receive alerts in a hierarchical manner, enabling them to respond more accurately to the most critical situations first.

4.5 Continual Learning-Based Behavior Rules Update

Despite categorizing abnormal behaviors under broad types, significant differences exist within each category that necessitate the accumulation of a substantial database for precise abnormal behavior recognition. Figure 10 shows the examples of different instances within the same abnormal behavior category (SelfHarm (Hit Head)). Figure 10(a) means how the behaviors within a given abnormal behavior can be further divided into classes 1-3 and share a common "Similar spacered as defined in Section 3. Figure 10(b) shows examples of different instances within an abnormal behavior class (SelfHarm (Hit Head)). They have variations in posture and execution, such as sitting vs standing and using one or both hands. Despite belonging to the same behavior category, these variations result in distinct feature vectors.

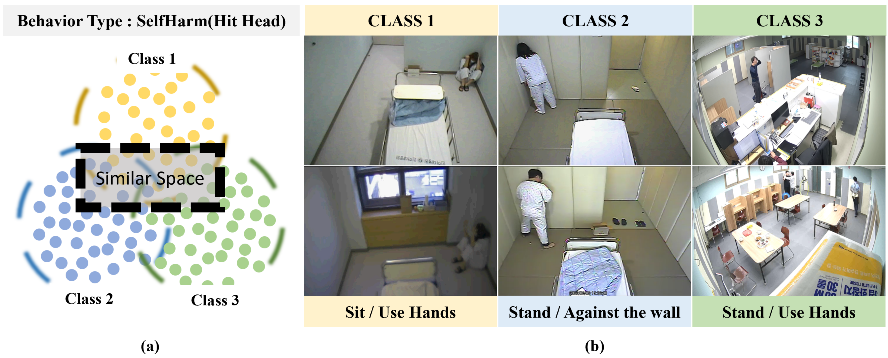


Figure 10: Characteristics of the abnormal behaviors. Even within a given type of behavior, leading to division under several classes owing to their significant variations, (a) several classes may share a common location, but exhibit distinct features. (b) Examples of classes within the SelfHarm (Hit Head) behavior type.

However, it is difficult to obtain sufficient amounts of abnormal behavior data for patients. To solve this problem, the proposed abnormal behavior recognition system adopted a continual learning method to acquire diverse data on abnormal behaviors. Using this approach, the proposed algorithm demonstrated the capability to learn from a continuous stream of data, allowing it to adapt to shifts in data distribution or accommodate new tasks. Importantly, this adaptive learning process was designed to minimize the loss of knowledge acquired from previous experiences.

The process of abnormal behaviors updates from the continuous data stream in CCTV environments is outlined as follows:

1. Assigning labels to frame durations based on behavior type.
2. Computing feature vectors representing differences between adjacent frames.
3. Performing clustering among feature vectors with high similarities for each behavior type.
4. Updating the personalized behavior rules and database.

Let's represent the target data of frame duration as $\mathbf{x} = \{x_0, x_1, \cdots, x_n\}$, where n corresponds to the frame number within the range $0 \leq n \leq N$. The initiation of an arbitrary frame duration $\mathbf{x}$ occurs when a detected human starts moving or falls within the initial definitions. Each frame duration $\mathbf{x}$ is assigned a label based on its behavior type, and concurrently, a feature vector $\mathbf{s_x}$ is computed for each $\mathbf{x}$. This feature vector $\mathbf{s_x}$ encapsulates the differences between adjacent frames within the frame duration $\mathbf{x}$. It is a $1 * 24$ dimensional vector comprising posture information, 13 joint distances, 6 joint angles, and other characteristics specific to each behavior type. To enhance the analysis, clustering is conducted among feature vectors exhibiting high similarity within each behavior type. Let C_{bi} denote the i^{th} class of a particular behavior type b, and μ_{bi} represent the centroid of cluster C_{bi}. The classification of $\mathbf{s_x}$ is carried out using a similarity metric called *ASIM*. *ASIM*, quantifying the similarity with each class red, is measured by the following equation, Equation (1):

$$ASIM(\mathbf{s_x}, C_{bi}) = \frac{\mathbf{s_x} \cdot \mu_{bi}}{|\mathbf{s_x}|\,|\mu_{bi}|}. \tag{1}$$

If a feature vector $\mathbf{s_x}$ is similar to behavior type b, it is recognized as a new vector of behavior type b, updating the centroid by averaging with this new vector. If a feature vector $\mathbf{s_x}$ is not similar to any class, a particular frame duration $\mathbf{x}$ is defined as a new class. This iterative process is applied continuously to incoming data streams, leading to the regular update of abnormal behavior rules and the associated database. Through these ongoing iterations, the availability of abnormal behavior data is enhanced, ultimately contributing to the continuous refinement and improvement of the accuracy of abnormal behavior recognition.

4.6 Abnormal behavior Alarm

Considering the main goal of minimizing false negatives within closed wards, it is imperative to address the potential occurrence of false alarms even after filtering through Skeleton Estimation and Rule-based Behavior Recognition. For this reason, the proposed alarm system was configured to trigger an alarm when a specific type of behavior was detected multiple times. To enhance the accuracy of alarms, the required number of detections of each behavior type had to be set dependent on the risk of behavior as shown in Table 5. For Falldown and Hit, a low threshold of the number of detections was needed since these behaviors were quite risky. On the contrary, SelfHarm and Caution had a lower risk of immediate injury. These behaviors may not lead to severe injuries in a short time. This is why they had higher thresholds than those of Falldown and Hit.

Table 5: Repeated detection counts of each behavior type for producing an alarm. The repeated detection counts of Falldown and Hit were set to lower values because of their high probabilities of sustaining injuries. Conversely, the repeated detection counts were set high for SelfHarm and Caution.

Behavior Type	**Count**	**Behavior Type**	**Count**
SelfHarm	10	Caution	20
Falldown	5	Hit	5

5 Database

In this paper, we not only defined abnormal behaviors that were previously undefined but also concurrently established a database for these behaviors. Recognizing the limited availability of databases for abnormal behaviors, particularly within hospital environments, the research team collaborated with Inha University Hospital to establish a database specifically focused on abnormal behaviors exhibited by patients. The collection of data strictly adhered to individual patient consent, and it is important to note that no personal information beyond gender was included in the databases.

For the initial database, data were obtained from Inha University Hospital's closed wards through CCTV, resulting in a total of 508 videos of 13 people. There were 9-10 videos available for each of the four major behavior types (SelfHarm, Falldown, Caution, and Hit) with each video lasting between 1 and 2 min. Notably, unlike typical behavior recognition datasets that have a single label for a video, the dataset provides labels for each frame, as illustrated in Figure 11. Professional nurses from Inha University Hospital with specialized medical knowledge provided these labels.

In addition to the initial dataset for defining initial abnormal behavior types, another database was constructed through continual learning from streaming data. Cameras were installed in two intensive care rooms within the psychiatric ward at Inha University Hospital for over 12 weeks. The tests of the proposed system were conducted and additional data were collected from 10 patients for the entire 12 weeks. It yielded approximately 980 videos, including 20-25 videos for each behavior type. The overall specifications of the final database, including the initial dataset, are summarized in Table 6. The dataset used for performance evaluation accounted for a total of 30 min per person, comprising approximately 1,500,000 frames.

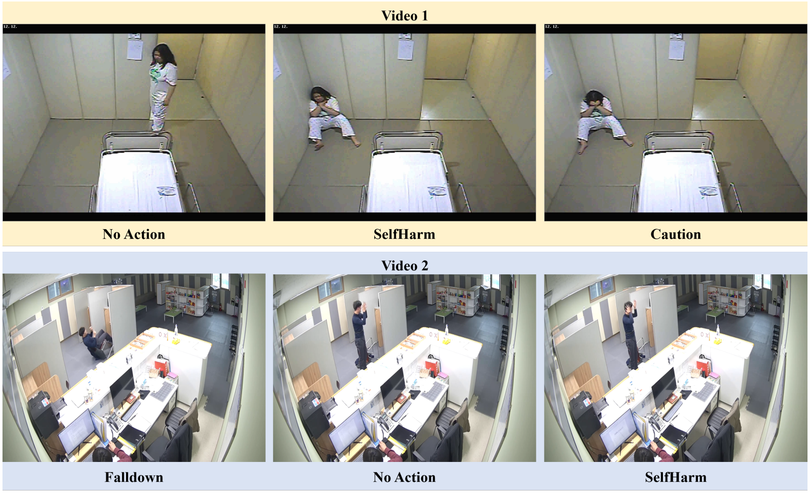


Figure 11: Examples from our dataset, where behavior labels are assigned at the frame level. The meticulous labeling process, guided by input from healthcare professionals, ensures a reliable foundation for training and evaluating the abnormal behavior recognition system.

Table 6: Final database setup via continual learning. All labels for the database were provided by professional nurses at Inha University Hospital to ensure the accuracy and reliability of the dataset.

Items	Value
# of Persons	23
# of Videos	1,492
# of Frames	2,546,300
Frame Rate	12.5 fps

6 Experiments

6.1 Experimental Setup

We conducted testing at Inha University Hospital over a total of 12 weeks. For the previous two weeks, we collected the initial database mentioned in Section 5, and for the remaining 10 weeks, we checked the performance of the proposed algorithm. For the test, we used real-time videos from two cameras in a private room and one in a common space at Inha University Hospital. For each camera, an GeForce RTX 2080 was used for the computation. Through a 10-week test, we checked the performance differences by gradually updating the behavior rules of the proposed algorithm based on data received through CCTV.

6.2 The Accuracy of Abnormal Behavior Recognition

The results of our proposed system are shown in Figure 12. All results are displayed on the streaming video. Therefore, medical staff can check mental patients anytime they want. Table 7 also shows an overview of the performance evaluation of the proposed abnormal behavior recognition alarm system. The evaluation shows the comparison of the abnormal behavior labels with the ground truth labels in a total of 312 videos. The confusion matrix depicts the recognition accuracy for five behavior types: SelfHarm, Caution, Hit, Falldown, and No Action. In conventional behavior recognition datasets, short videos typically have one behavior label. However, our abnormal behavior dataset consists of single-long videos with multiple behavior labels each. Therefore, measuring the recognition accuracy for the "No Action" category is crucial. As illustrated in Table 7, the average accuracy for abnormal behavior recognition reached 95.5%. Notably, the system achieved a high accuracy of 97.3% for correctly classifying "No Action" when it indeed corresponds to no abnormal behaviors. This accuracy is critical for efficient management by medical staff, ensuring precise alarms for abnormal behaviors and facilitating appropriate responses timely.

Table 7: Abnormal behavior recognition accuracy of the proposed system. The left column represents the categories of the estimated results, and the top row represents the ground truth labels.

	No Action	**SelfHarm**	**Caution**	**Hit**	**Falldown**
No Action	0.987	0.032	0.029	0.083	0.110
SelfHarm	0.004	0.921	0.001	0	0.002
Caution	0.007	0.045	0.970	0	0
Hit	0.001	0	0	0.917	0
Falldown	0.001	0.002	0	0	0.888

The accuracies for SelfHarm and Caution were also commendable, reaching 92.1% and 97.0%, respectively. Despite these behavior types involving relatively lower motions than others, such as Hit Head and Punch, the system captured limb movements effectively. However, Hit and Falldown exhibited slightly lower accuracies at 91.7% and 88.8%, respectively, compared to the other two abnormal behavior types. An analysis revealed that instances of Hit and Falldown often had very short prelude symptoms, leading to the initial frames being categorized as No Action. Additionally, the short durations of Hit and Falldown impacted accuracy when measured relative to the entire duration of the behavior. Further discussion on behavior recognition durations is provided in Section 6.5.

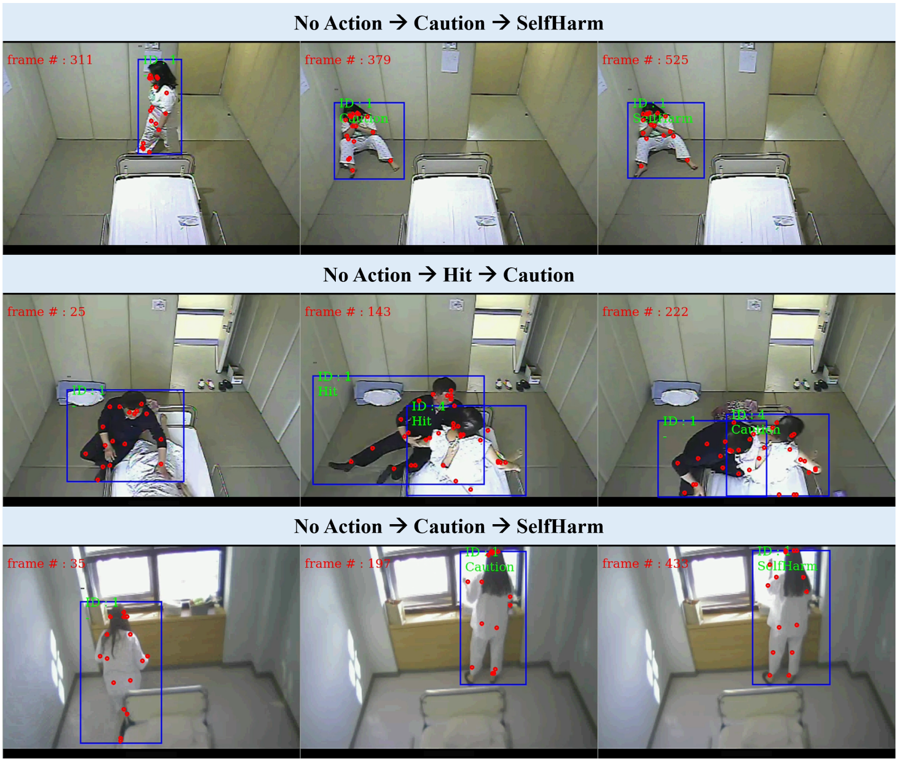


Figure 12: Results of our proposed system. The results in each row are generated from a single video and show multiple anomalous behavior labels. The results of each abnormal action are displayed in the image, and in the case of No Action, they are not displayed.

6.3 The comparison with other methods

To conduct a comprehensive performance evaluation, we compared the abnormal behavior recognition accuracy of our method with four existing algorithms [30, 7, 8, 22]. Due to the limited quantity of the database for training deep learning models extensively, pre-training was performed using the databases employed in each respective paper, excluding red. The results are summarized in Table 8, showing that our method achieved significantly higher abnormal behavior recognition accuracy compared to other algorithms.

The superiority of the proposed method is higher accuracy in recognizing "No Action" compared to other algorithms, where the recognition accuracy for this behavior was generally low. This observation is evident in the consistently lower results for other algorithms when "No Action" is excluded from the analysis. It is important to note that many behavior recognition algorithms

Table 8: Abnormal behavior recognition accuracy of each method.

Model	Pre-training	w/o No Action	w No Action
Yan *et al.* [30]	No	65.92%	69.32 %
Girdhar *et al.* [7]	Yes	64.22%	65.67%
Habib *et al.* [8]	Yes	72.01%	74.41%
Sato *et al.* [22]	Yes	78.03%	80.26%
Proposed method	No	**93.66%**	**92.4%**

typically predict behavior labels based on a fixed length of video, which differs from our proposed system where behavior labels are predicted for each frame. Due to this distinction, the recognition accuracy of other algorithms may be lower. It means that our proposed method prioritizes the accuracy of determining the presence or absence of behavior over the accuracy of behavior recognition.

6.4 False Positive Rate of Abnormal Behavior Recognition

In Section 6.2, we showed details on the recognition accuracy of each abnormal behavior type. However, in this paper, the main goal of the proposed system is to diminish the probability of false alarms which can be caused inefficiently. Therefore, the performance of the proposed system is considered through precision and recall values. Precision and recall are calculated as follows:

$$(Precision) = \frac{TP}{TP + FP},$$
$$(Recall) = \frac{TP}{TP + FN},$$

where TP is true positive, FP is false positive, and FN is false negative. Precision signifies the proportion of true positives among all instances classified as true by the model, while recall represents the proportion of actual true instances that were predicted as true by the model.

The proposed system achieved a precision of **0.9761** and recall of **0.9579**, both of which were indicating high values. Furthermore, the false negative rate was **1.51%**, an exceptionally low value. Also, the false positive rate was **0.84%**. These results demonstrate that the proposed method can effectively minimize inefficient calls to medical staff.

6.5 Recognition Time

The proposed alarm system not only strives to achieve high recognition accuracy and low false positive rate but also aims for fast recognition times. In a closed ward, predicting a patient's next behavior is exceptionally challenging, which makes it crucial to quickly intervene when an abnormal behavior observed. Table 9 shows the recognition time for each abnormal behavior type, and the recognition delay values are aligned closely with the recognition accuracy results. For behavior types with relatively longer lead-up symptoms, such as SelfHarm and Caution, the recognition delays were remarkably short, averaging 1.667 frames (0.133s) and 1.012 frames (0.081s), respectively. Conversely, for Falldown and Hit, which exhibit shorter lead-up symptoms, the recognition delays were slightly longer, averaging 4.333 frames (0.347s) and 3.028 frames (0.242s), respectively. These differences in delay is attributed to the shorter lead-up symptoms in Falldown and Hit, contributing to the slightly lower recognition accuracies for these behavior types.

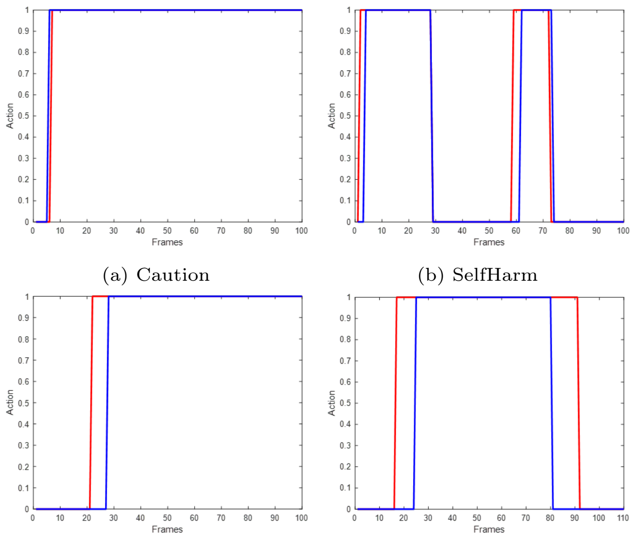
Table 9: Recognition time for each abnormal behavior type.

Behavior Type	**Frames**	**Time(s)**
Caution	1.012	0.081
SelfHarm	1.667	0.133
Falldown	4.333	0.347
Hit	3.028	0.242
Average	2.510	0.201

Figure 13 visually shows the abnormal behavior recognition delay times. Examples of the recognition delays for Caution and SelfHarm are shown in Figures 13(a) and 13(b), showing relatively quick recognition with differences of 1 and 2 frames compared to the ground truth, respectively. In contrast, Figures 13(c) and 13(d) display examples of recognition delay for Falldown and Hit, where the differences between the ground truth and estimated recognition were more pronounced with the delays of 3 and 5 frames, respectively, consistent with the results in Table 9. These slightly larger delays were attributed to the shorter lead-up symptoms for Falldown and Hit.

6.6 Ablation Study

This section presents comparisons on the abnormal behavior recognition performances with and without the continual learning block to assess the effectiveness of continual learning-based behavior rule updates. Table 10 shows the accuracies for the abnormal behavior types and average accuracies in the presence and absence of the continual learning block. The results indicate that the recognition accuracies for all five behaviors, including "No Action", are higher when the continual learning block exists.



(a) Caution (b) SelfHarm (c) Falldown (d) Hit

Figure 13: Examples of recognizing each abnormal behavior type. Here, red represents the ground truth and blue represents the results of abnormal behavior recognition, with 1 indicating detection and 0 indicating no detection for each abnormal behavior type.

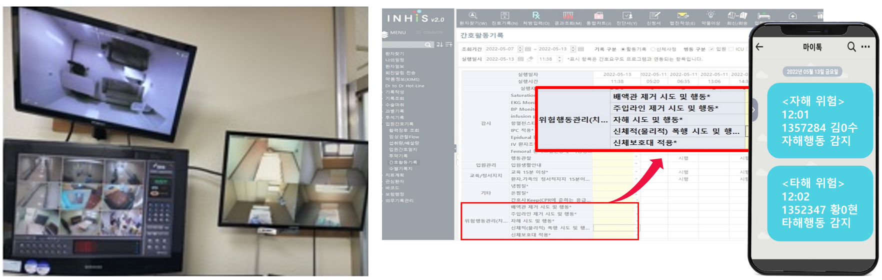
Table 10: Ablation study on the continual learning block. With the continual learning block, the abnormal behavior recognition performance is better than without this block.

	No Action	SelfHarm	Caution	Hit	Falldown	Avg.
w/o block	0.973	0.905	0.957	0.850	0.839	**0.904**
w block	0.987	0.921	0.970	0.917	0.888	**0.936**

The continual learning block facilitates the continual updating of the behavior rule whenever a new form of behavior is observed, leading to more detailed definitions than the initially established behavior rules. As a result, higher recognition accuracies are achieved for all abnormal behavior types. Notably, the recognition accuracies improve by over 5% for Hit and Falldown which exhibit wide variations, making it challenging to define them with simple features. Therefore, continual updates to the behavior rules across various classifications are essential for higher recognition accuracies. Through the continual learning block, the system not only updates personalized behavior rules but also enhances the database. This results in an effective system that simultaneously improves model performance and acquires a more comprehensive yet privacy-preserving hospital dataset.

6.7 Test System at a Hospital

The proposed personalized abnormal behavior recognition alarm system was tested at Inha University Hospital for approximately three months. Figure 14 shows the real system that medical staff used at Inha University. They used this system to check the alarm from our proposed method. Through the test, we analyzed how much abnormal behavior recognition performance has changed through this period. Also, we checked the impact on the working environment of the hospital staff.



(a) (b)

Figure 14: Real system in Inha University. (a) the real setting in Inha University, (b) the hospital's system for accepting alarms.

Figure 15 shows the results of real test in hospital. Each result shows how quickly the medical staff was able to respond when abnormal behaviors occurred. As a result, they arrived at the patient within 8s on average. It was possible since behavior recognition performance continued to improve during monitoring. This fact is able to be checked in Figure 16. It shows the average of abnormal behavior recognition accuracy for the 10 patients over each week. In the initial two weeks, the accuracy was not calculated since it collected the patient's data. From the third to the 12^{th} week, individual abnormal behavior accuracy for each patient was monitored. This upward trend suggests that as more meaningful data is obtained, the behavior rules are updated more accurately.

Not only technical performance but also practical usability in a real hospital is important for showing the superiority of the proposed system. Figure 17 shows the survey of 25 staff members from a psychiatric ward at Inha University Hospital. As shown in Figure 17(a), of the 25 staff members, 21 (84%) responded positively, stating that the proposed system had a beneficial effect on improving their working environment. Among the 21 staff members who responded positively, 14 expressed satisfaction with the system's ability to respond quickly to situations, as depicted in Figure 17(b). In addition, five staff members appreciated the system's capacity to share responsibilities with

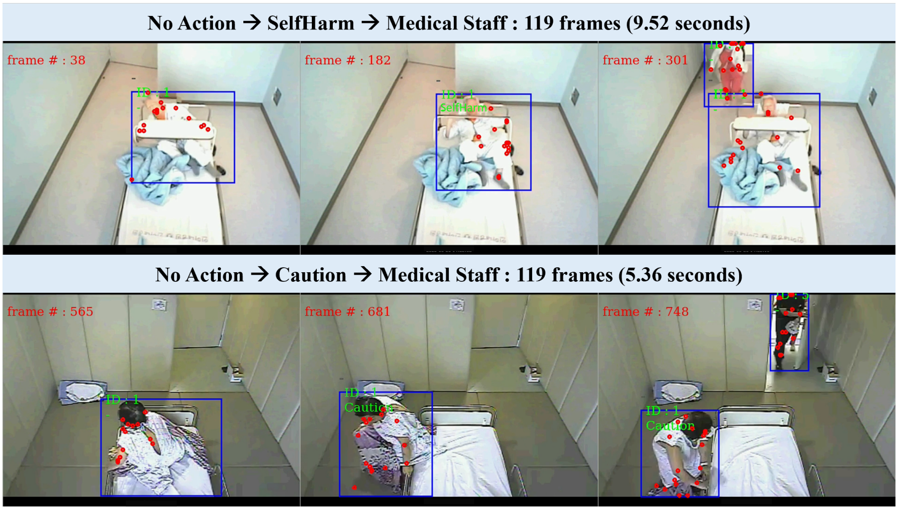


Figure 15: Results of testing in hospital. Each result shows how quickly the medical staff was able to respond when SelfHarm and Caution occurred. In the case of SelfHarm, medical staff arrived about 9 seconds after the abnormal behavior occurred, and in the case of Caution, about 5s later.

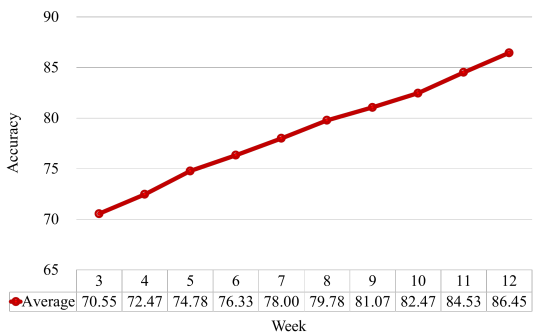


Figure 16: Average results of abnormal behavior recognition over 12 weeks, excluding the initial 2 weeks for data collection. The recognition accuracy gradually to increased from the 3^{rd} week onward.

other tasks, while the remaining two were unsure. These survey responses, combined with the technical performance results, affirm that the proposed personalized abnormal behavior recognition alarm system is effective for use in a real hospital environment.

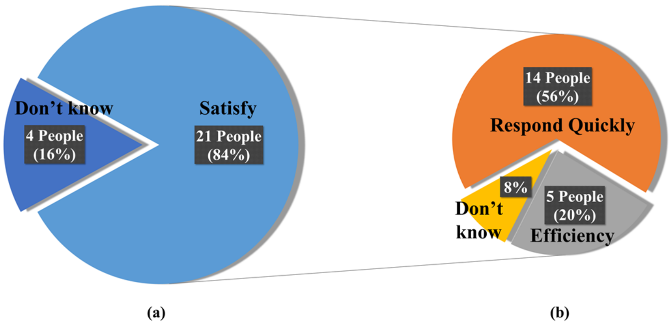


Figure 17: Survey results of target employees working at Inha University Hospital. About 84% of the medical staff were satisfied with the proposed alarm system of which 56% expressed satisfaction because of the fast responses.

7 Conclusion

A personalized abnormal behavior recognition alarm system for close monitoring of mental patients was successfully designed and implemented to analyze individual behavioral characteristics and facilitate personalized abnormal behavior recognition in a closed ward. The algorithm promptly transmits detected information in the form of an alarm through the hospital alert system, enabling swift responses from medical staff. Through continual learning, the algorithm acquires meaningful individual patient data from streaming data to continuously update behavior rules so as to enhance ongoing abnormal behavior recognition performance. This sustainable and personalized abnormal behavior recognition alarm system not only facilitates rapid responses from the medical staff but also addresses challenges related to a shortage of medical personnel within psychiatric hospitals. The effectiveness of this system was demonstrated through a three-month-long test in a real hospital setting.

Although this study focused mainly on commonly observed abnormal behaviors among psychiatric patients, the diverse behavioral patterns among people with mental disorders suggest the potential for personalized definitions of abnormal behaviors for future work. By defining unique behavioral patterns for each individual and training the proposed personalized abnormal behavior recognition system based on these definitions, the system could evolve into a tool for safeguarding the daily lives of individuals with mental illnesses, not only within psychiatric wards but also in their everyday residences. This extension would broaden the system's utility to protect the daily routines of caregivers and family members.

Acknowledgements

This work was supported by the Grant of the Korea Health Technology Research and Development Project through the Korea Health Industry Development Institute (KHIDI), Ministry of Health and Welfare, Republic of Korea, under Grant HI22C1884.

References

[1] D. H. Barlow, V. M. Durand, and S. G. Hofmann, "Abnormal psychology: An integrative approach", in, USA: Cengage learning, 2018.

[2] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection", *arXiv preprint arXiv:2004.10934*, 2020, DOI: 10.48550/arXiv.2004.10934.

[3] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values", *Scientific reports*, 8(1), 2018, 6085, DOI: 10.1038/s41598-018-24271-9.

[4] C. Chen, R. Jafari, and N. Kehtarnavaz, "UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor", in *2015 IEEE International conference on image processing (ICIP)*, IEEE, 2015, 168–72.

[5] *DeepStream SDK | NVIDIA Developer*, https://developer.nvidia.com/deepstream-sdk.

[6] M. Elhoseny, "Multi-object detection and tracking (MODT) machine learning model for real-time video surveillance systems", *Circuits, Systems, and Signal Processing*, 39, 2020, 611–30, DOI: 10.1007/s00034-019-01234-7.

[7] R. Girdhar, J. Carreira, C. Doersch, and A. Zisserman, "Video action transformer network", in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, 244–53, DOI: 10.1109/CVPR.2019.00033.

[8] S. Habib, A. Hussain, W. Albattah, M. Islam, S. Khan, R. U. Khan, and K. Khan, "Abnormal activity recognition from surveillance videos using convolutional neural network", *Sensors*, 21(24), 2021, 8291, DOI: 10.3390/s21248291.

[9] A. M. Kring and S. L. Johnson, "Abnormal psychology: The science and treatment of psychological disorders", in, USA: John Wiley & Sons, 2022.

[10] B. Kwon and T. Kim, "Toward an online continual learning architecture for intrusion detection of video surveillance", *IEEE Access*, 10, 2022, 89732–44.

[11] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "LSTM-based encoder-decoder for multi-sensor anomaly detection", *arXiv preprint arXiv:1607.00148*, 2016, DOI: 10.48550/arXiv.1607.00148.

[12] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine learning for anomaly detection: A systematic review", *Ieee Access*, 9, 2021, 78658–700, DOI: 10.1109/ACCESS.2021.3083060.

[13] T.-N. Nguyen and J. Meunier, "Anomaly detection in video sequence with appearance-motion correspondence", in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, 1273–83, DOI: 10.1109/ICCV.2019.00136.

[14] W. H. Organization *et al.*, "Mental health atlas 2020: review of the Eastern Mediterranean Region", in, Switzerland: World Health Organization. Regional Office for the Eastern Mediterranean, 2022.

[15] W. H. Organization *et al.*, "World health statistics 2020", 2020.

[16] A. Pramanik, S. Sarkar, and J. Maiti, "A real-time video surveillance system for traffic pre-events detection", *Accident Analysis & Prevention*, 154, 2021, 106019, DOI: 10.1016/j.aap.2021.106019.

[17] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement", *arXiv preprint arXiv:1804.02767*, 2018.

[18] M. Reinert, D. Fritze, and T. Nguyen, "The state of mental health in America 2022", 2021.

[19] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep one-class classification", in *International conference on machine learning*, PMLR, 2018, 4393–402.

[20] A. B. Sargano, X. Gu, P. Angelov, and Z. Habib, "Human action recognition using deep rule-based classifier", *Multimedia Tools and Applications*, 79(41), 2020, 30653–67.

[21] F. Sato, R. Hachiuma, and T. Sekii, "Prompt-Guided Zero-Shot Anomaly Action Recognition using Pretrained Deep Skeleton Features", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, 6471–80, DOI: 10.1109/CVPR52729.2023.00626.

[22] F. Sato, R. Hachiuma, and T. Sekii, "Prompt-Guided Zero-Shot Anomaly Action Recognition using Pretrained Deep Skeleton Features", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, 6471–80, DOI: 10.1109/CVPR52729.2023.00626.

[23] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery", in *International conference on information processing in medical imaging*, Springer, 2017, 146–57, DOI: 10.1007/978-3-319-59050-9_12.

[24] A. Shahroudy, J. Liu, T.-T. Ng, and G. Wang, "Ntu rgb+ d: A large scale dataset for 3d human activity analysis", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, 1010–9.

[25] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation", in *CVPR*, 2019, DOI: 10.1109/CVPR.2019.00584.

[26] N. Varshney, B. Bakariya, A. K. S. Kushwaha, and M. Khare, "Rule-based multi-view human activity recognition system in real time using skeleton data from RGB-D sensor", *Soft Computing*, 2023, 1–17.

[27] K. K. Verma, B. M. Singh, and A. Dixit, "A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system", *International Journal of Information Technology*, 14(1), 2022, 397–410, DOI: 10.1007/s41870-019-00364-0.

[28] L. Wang, X. Zhang, H. Su, and J. Zhu, "A comprehensive survey of continual learning: Theory, method and application", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, DOI: 10.48550/arXiv.2302.00487.

[29] B. Xiao, H. Wu, and Y. Wei, "Simple Baselines for Human Pose Estimation and Tracking", in *European Conference on Computer Vision (ECCV)*, 2018, DOI: DOI:10.1007/978-3-030-01231-1_29.

[30] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition", in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32, No. 1, 2018, DOI: 10.48550/arXiv.1801.07455.

[31] Y. Zhang, W. Liang, X. Yuan, S. Zhang, G. Yang, and Z. Zeng, "Deep Learning Based Abnormal Behavior Detection for Elderly Healthcare Using Consumer Network Cameras", *IEEE Transactions on Consumer Electronics*, 2023.